

**IN THE UNITED STATES DISTRICT COURT FOR
THE NORTHERN DISTRICT OF GEORGIA
ATLANTA DIVISION**

DONNA CURLING, et al.

Plaintiff,

vs.

BRAD RAFFENSPERGER, et al.

Defendant.

**CIVIL ACTION FILE NO.:
1:17-cv-2989-AT**

DECLARATION OF PHILIP B. STARK

PHILIP B. STARK hereby declares as follows:

This statement supplements my declarations of September 9, 2018; September 30, 2018; October 22, 2019; December 16, 2019; August 23, 2020; August 31, 2020; September 13, 2020; August 2, 2021; and January 11, 2022 (augmented March 9, 2022), and December 5, 2022. I stand by everything in my previous declarations and incorporate them by reference.

1. I have personal knowledge of all facts stated in this declaration, and if called to testify, I could and would testify competently thereto.

2. Exhibit 1 is copy of a law journal paper I co-authored with Dr. Andrew Appel, entitled *Evidence-Based Elections: Create a Meaningful Paper Trail, Then Audit*. It appeared in 2020 in *Georgetown Technology Law Review*, 4,

523. It explains why the paper printout from Ballot Marking Devices (“BMDs”) is not a trustworthy source record for purposes of tabulating or auditing public elections, and explains some of the steps needed to establish whether a paper trail of (hand-marked) ballots is trustworthy prior to conducting a risk-limiting audit.

3. Exhibit 2 is the final, published, peer-reviewed version of a paper I co-authored, *They May Look and Look, Yet Not See: BMDs Cannot be Tested Adequately*, which appeared in 2022 in the proceedings of E-Vote-ID 2022, LNCS 13553, pp. 122–138 DOI10.1007/978-3-031-15911-4_8. An earlier version of this article was entitled *There is no Reliable Way to Detect Hacked Ballot-Marking Devices*, which is attached as Exhibit 3. I cite this paper in several declarations. The paper shows that no feasible amount of logic and accuracy testing, parallel testing, or “passive” testing can ensure that BMD misbehavior does not change contest outcomes.

4. Research conducted by University of Georgia faculty at the behest of the Georgia Secretary of State found that voter verification rates are low and voters who do look at their printed BMD ballots generally look for very brief periods. This research, *Georgia Voter Verification Study Draft Version 1.0*, reported in January 2021, is attached as Exhibit 4. I cite this report in my declarations. The copy of the report was obtained from the Atlanta Journal Constitution published

article.¹ The reporter states that “The Georgia study is the largest known examination of voter behavior on ballot-marking devices, with over 4,000 voters observed in 39 precincts.” I am aware of no information to the contrary.

5. I am not aware of any research that is inconsistent with the research cited in Exhibit 4, nor any research that finds that voters can or do verify their BMD ballots well enough to ensure that BMD misbehavior does not alter contest outcomes.

6. I co-authored a September 8, 2022, letter to Georgia state election officials concerning the breach of the Georgia Voting System that was initiated in Coffee County. A true and correct copy is attached as Exhibit 5. On page 2, the letter explains why “Georgia’s elections have a higher risk of compromise due to the reliance on computerized Ballot-Marking Devices (BMDs) to record vote selections for in-person voting.”

7. Signers of the letter include prominent members of the community of scientists who are experts in voting systems. Several of the co-signers are current or former members of the Verified Voting Board of Directors or Board of Advisors, such as Dr. Duncan Buell, Dr. Richard DeMillo, Dr. Larry Diamond,

¹ AJC: *Under Half of Georgia voter checked their paper ballots, study shows.*
<https://www.ajc.com/politics/under-half-of-georgia-voters-checked-their-paper-ballots-study-shows/6HSVHHFOBRBDPODRZXLIBTUS64/>

Lowell Finley, Dr. David Jefferson, Dr. Douglas Jones, Mark Ritchie, Dr. John Savage, and Dr. Kevin Skoglund, most whom I know personally through Verified Voting, election integrity conferences, or other election integrity activities.

Verified Voting's Board current members are listed on their website at

<https://verifiedvoting.org/team/#advisors>

8. I co-authored a September 2, 2021, letter to California Secretary of State Dr. Shirley Weber warning of the security issues related to the release of Dominion software from Antrim County, Michigan and Mesa County, Colorado. A true and correct copy of the letter is attached as Exhibit 6. The letter explains the increased security risk to jurisdictions using the Dominion BMD machines. The co-signers were well-respected members of the community of voting system experts; several of them are computer scientists.

9. I co-authored a December 12, 2022, letter to the U.S. Department of Justice, the Federal Bureau of Investigation and Cybersecurity & Infrastructure Agency (CISA) seeking federal investigations of the Georgia voting systems breach. A true and correct copy of the letter is attached as Exhibit 7. The letter explains the elevated risk of election manipulation because of Georgia's universal use of BMDs. (page 11)

10. Joining me in submitting this letter were highly respected members of the scientific and legal community with voting systems expertise. (pages 12-14)

11. OSET Institute, a non-profit organization with emphasis on election technology, issued a statement regarding the risks to voting systems in the wake of the breach of the Georgia system through Coffee County. A true and correct copy of the statement is attached as Exhibit 8 as published and circulated. It explains that the risk is heightened because of Georgia's use of BMDs. (footnote 7 page 5)

12. I serve on the OSET Institute Board of Advisors, which includes several other voting system experts: Dr. Barbara Simons (Chair of the Board of Verified Voting), Dr. Richard DeMillo (Georgia Tech), and Dr. Stephanie Singer (Portland State). Gregory Miller, John Sebes, and Edward Perez, John Gage, and seven others serve on the OSET Board of Directors.

13. Dr. Andrew Appel incorporated some of the above-referenced research in which I was involved in his July 21, 2022 article, "Magical Thinking about Ballot-Marking-Device contingency plans." That article is attached as Exhibit 9 (downloaded January 15, 2023).

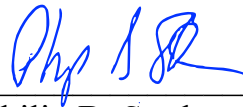
14. My December 5, 2022, Declaration (Doc. 1569-44 ¶24, and footnote 15) references Dr. Halderman's and Dr. Springall's "DVSTest" research. Exhibit 10 is a true and correct copy of the DVSTest research publication, downloaded February 1, 2023.

15. My March 9, 2022, Declaration references the Secretary of State's press release concerning the November 2020 post election audit. (Doc. 1569-42,

footnote 4) Exhibit 11 is a copy of that press release downloaded from the Secretary of State's website on January 21, 2023.

I declare under penalty of perjury, in accordance with 28 U.S.C. § 1746, that the foregoing is true and correct.

Executed on this date, February 5, 2023.



Philip B. Stark

EVIDENCE-BASED ELECTIONS: CREATE A MEANINGFUL PAPER TRAIL, THEN AUDIT

Andrew W. Appel* & Philip B. Stark**

CITE AS: 4 GEO. L. TECH. REV. 523 (2020)

TABLE OF CONTENTS

I.	INTRODUCTION	523
II.	VOTER-VERIFIED PAPER BALLOTS	525
	A. Hand-Marked Paper Ballots (Optical Scan)	525
	B. Direct-Recording Electronic (DRE) Machines	526
	C. Voter-Verifiable Paper Audit Trail (VVPAT)	526
	D. Ballot-Marking Devices (BMDs)	527
	E. All-In-One BMDs	528
	F. Internet Voting	529
	G. Software Independence, Contestability, Defensibility	529
III.	RISK-LIMITING AUDITS	530
IV.	COMPLIANCE AUDITS	532
V.	EFFICIENT RISK-LIMITING AUDITS	534
VI.	RESOURCES FOR RISK-LIMITING AUDITS	536
	A. Audit the Digital Images?	537
VII.	PRINCIPLES FOR ELECTION INTEGRITY LEGISLATION	537
VIII.	CONCLUSIONS	540

I. INTRODUCTION

There is no perfect, infallible way to count votes. All methods—including optical scan, touchscreen, and hand counting—are subject to errors, procedural lapses, and deliberate manipulation. Almost all U.S. jurisdictions count their votes using computer-based technology, such as touchscreens and optical-scan machines. Computer-based methods are subject to “hacking,” that is, the replacement of legitimate vote-counting software with a computer

* Eugene Higgins Professor of Computer Science, Princeton University.

** Professor of Statistics, University of California, Berkeley.

program that changes (some fraction of) the votes in favor of the hacker's preferred candidate(s). Hacking can be performed remotely (even if the machines are supposedly "never connected to the Internet") and it is very difficult to detect. Voters and election administrators see nothing out of the ordinary.

The vulnerability of computers to hacking is well understood. Modern computer systems, including voting machines, have many layers of software, comprising millions of lines of computer code; there are thousands of bugs in that code.¹ Some of those bugs are security vulnerabilities that permit attackers to modify or replace the software in the upper layers, so we can never be sure that the legitimate vote-counting software or the vote-marking user interface is actually the software running on election day.²

One might think, "our voting machines are never connected to the Internet, so hackers cannot get to them." But all voting machines need to be programmed for each new election: They need a "ballot-definition file" with the contests and candidate names for each election, and lists of the contests different voters are eligible to vote in. This programming is typically done via removable media such as a USB thumb drive or a memory card. Vote-stealing malware can piggyback on removable media and infect voting machines—even machines with no network connection.³

There is a way to count votes by computer and still achieve trustworthy election outcomes. A trustworthy paper trail of voter selections can be used to check, or correct, the electoral outcomes of the contests in an election.

¹ Estimates of software defect rates range from one per thousand lines of code (in high quality commercial products) down to 0.1 per thousand lines of code in extremely high-quality products (this is at the 90th percentile for the software industry). MEL LLAGUNO, SYNOPSYS, INC., 2017 COVERITY SCAN REPORT: OPEN-SOURCE SOFTWARE—THE ROAD AHEAD 16 (2017), <https://www.synopsys.com/content/dam/synopsys/sig-assets/reports/SCAN-Report-2017.pdf> [<https://perma.cc/H8L3-SADH>]. Modern voting machines contain software components such as an operating system (e.g., Windows 7 is fifty million lines of code; or Linux is twenty-seven million lines). USB drivers (common on voting machines) are quite large software components and are riddled with insecurities. Dave Tian et al., *SoK: "Plug & Pray" Today – Understanding USB Insecurity in Versions 1 through C*, IEEE SYMP. SECURITY AND PRIVACY (2018). Therefore, we can expect 100 to 1000 bugs per million lines of code; some small portion of these are "exploitable vulnerabilities," that is, an adversary can exploit them to take over the computer and install fraudulent software. A software-based product such as a voting machine can be expected to contain, at any given time, one or more exploitable security vulnerabilities.

² NAT'L ACADS. SCIS., ENG'G, & MED., SECURING THE VOTE: PROTECTING AMERICAN DEMOCRACY 89–91 (2018).

³ Ariel J. Feldman et al., *Security Analysis of The Diebold Accuvote-TS Voting Machine*, Proc. 2007 USENIX/ACCURATE ELECTRONIC VOTING TECH. WORKSHOP (2007).

“Electoral outcome” means the winning candidates or positions,⁴ not an exact numerical tally.

The principle of “evidence-based elections”⁵ is that local election officials should not only find the true winner(s) of an election, but they should also provide the electorate convincing evidence that they did. Generally, that means that eligible voters must have had the opportunity to vote, the election must have used voter-verified paper ballots, there must be convincing evidence that those ballots were kept inviolate through the audit, and the reported outcomes must be checked against the paper trail by suitable audits or hand counts.

To have affirmative evidence that reported outcomes are correct requires conducting elections using an auditable voting system, then auditing the results appropriately. First, we discuss audibility, or the creation of a trustworthy paper trail. Second, we discuss auditing—the method for efficiently assessing whether the computer-reported election outcomes are correct, based on the paper trail.

II. VOTER-VERIFIED PAPER BALLOTS

Society wants evidence that election outcomes are correct (e.g., the candidate actually selected by the voters wins the election), even if the computers have been hacked. The only known practical way to have trustworthy ballots to audit, even if the computer software has been hacked, is to have paper ballots, marked with the voters’ choices, that are manually interpretable, accountable, auditable, and re-countable.

A. Hand-Marked Paper Ballots (Optical Scan)

The traditional method of creating this paper trail (since about 1890 in the U.S.) is the use of a preprinted ballot form that lists, for each contest, the names of the candidates. Alongside each candidate is a target (square, oval, etc.) in which the voter indicates a vote. In recent decades, as such ballots are counted by optical scanners, the voter is asked to fill in an oval or complete an arrow to indicate selections. This is a *hand-marked paper ballot*.

With a hand-marked paper ballot, the marks on the ballot necessarily reflect what the voter did, and we can have reasonable assurance that the human-readable mark on the ballot is for the candidate actually intended by the voter. This assurance increases if the ballot follows standard best-practice ballot-design guidelines, such as those published by the U.S. Election

⁴ Or, for instance, whether there is a runoff.

⁵ Philip B. Stark & David Wagner, *Evidence Based Elections*, 10 IEEE SECURITY & PRIVACY 33–41 (2012) (introducing the concept of evidence-based elections).

Assistance Commission.⁶ Voters are more likely to overlook certain contests on the ballot, to overvote, to undervote, and to make other mistakes if the ballots do not follow these design guidelines.

Hand-marked paper ballots can be quite accurate: in the 2008 Minnesota election for U.S. Senator, of 2.4 million votes cast, only 0.01% (1 in 10,000) was so ambiguous that the State Canvassing Board could not interpret it, and the optical-scan voting machines agreed with the hand-recount totals with an accuracy of 99.99%.⁷

B. Direct-Recording Electronic (DRE) Machines

Direct-recording electronic voting machines have a user interface (typically a touchscreen) and an internal computer. Voters indicate their votes on the touchscreen, and the computer program interprets those indications to add votes to counters in its memory. At the close of the polls, the computer outputs the results by printing them on paper and saving results to a removable-media cartridge.

With a DRE, the record of the vote does not necessarily reflect what the voter did. If the DRE is “hacked,” that is, if fraudulent software is installed, then the fraudulent computer program can report arbitrary fraudulent votes. There is no effective paper trail. Results may be printed on paper at the close of polls, but this paper trail *starts* only when the computer program is reporting the totals. That printout can be effective in auditing the *aggregation* of votes from different precincts, but it cannot serve as a check on the computer program in the voting machine. This is a fatal flaw of paperless DRE voting machines.

In the early 21st century, many states used DRE voting machines. But, because of the widespread recognition of this fatal flaw, only a handful of states use paperless DRE voting machines, and many of those states are transitioning to technologies that have a paper trail starting from the individual voter’s ballot.

C. Voter-Verifiable Paper Audit Trail (VVPAT)

In the 2000s, it was thought that a good solution to the problem of DREs was a voter-verifiable paper audit trail. For a DRE with VVPAT, the

⁶ U.S. ELECTION ASSISTANCE COMM’N, EFFECTIVE DESIGNS FOR THE ADMINISTRATION OF FEDERAL ELECTIONS (2007), https://www.aiga.org/globalassets/migrated-pdfs/eac_effective_election_design [<https://perma.cc/G7H3-RDUL>].

⁷ Andrew Appel, *Optical-Scan Voting Extremely Accurate in Minnesota*, FREEDOM TO TINKER (Jan. 21, 2009), <https://freedom-to-tinker.com/2009/01/21/optical-scan-voting-extremely-accurate-minnesota/> [<https://perma.cc/4FWX-6GUJ>].

voter indicates choices on a DRE touchscreen. Then, the DRE prints the voter's selections on paper, behind glass. The voter inspects ("verifies") the VVPAT; and the VVPAT serves as the ballot of record in recounts or audits.

As we discuss below, VVPAT is not an adequate solution: in practice, the vast majority of voters do not verify the paper printout—it is "voter verifiable" but not "voter verified"; and the few who do inspect the VVPAT cannot safeguard the votes of their fellow voters who do not.

D. Ballot-Marking Devices (BMDs)

Ballot-marking devices have a user interface (typically a touchscreen) on which voters indicate their selections; then the BMD prints a paper ballot that will be optically scanned. There are many variations of this technology: the paper ballot may have only human-readable marks or the votes may also be encoded in barcodes. The paper ballot might print a summary of the voter's selections, or also contests the voter skipped. The paper ballot may be displayed under glass; it may be ejected for the voter to hold and inspect before feeding back into a slot for scanning; or it may be ejected for the voter to carry to a separate optical scanner.

None of these designs is trustworthy. Just like DRE VVPATs, a BMD vote record might not reflect what the voter did. BMDs print out a paper ballot that is, in principle, *voter verifiable*, but is not, in practice, *voter verified*. In a study of voters using BMDs in a 2018 election in Tennessee, DeMillo et al. found that 47% of voters did not inspect their BMD-printed ballots at all; the other 53% looked at their paper ballot for an average of 3.9 seconds, not nearly long enough to check that the printout matched what they indicated on the touchscreen for all 18 contests on the ballot.⁸ In a controlled experiment with real voters (but not in a real election), Bernhard et al. found that when the BMD deliberately mis-recorded one vote on each ballot, only 7% of the voters noticed.⁹

If a BMD is hacked and systematically steals 5% of the votes in one contest and only 7% of voters inspect their ballots carefully enough to notice, then the effective rate of vote-theft is $5\% \times 93\%$, or 4.65%; this is enough to change the outcome of a moderately close election. The same analysis applies to a DRE+VVPAT system.

One might think: "not everyone needs to carefully verify their ballots;" if only 7% of voters carefully inspect their ballots, they can serve as a kind of

⁸ Richard DeMillo et al., *What Voters Are Asked to Verify Affects Ballot Verification: A Quantitative Analysis of Voters' Memories of Their Ballots* (last revised Apr. 13, 2019) (unpublished manuscript) <https://ssrn.com/abstract=3292208> [<https://perma.cc/6VYR-FFZJ>].

⁹ Matthew Bernhard et al., *Can Voters Detect Malicious Manipulation of Ballot Marking Devices?*, 41 IEEE SYMP. SECURITY & PRIVACY (2020).

“random audit” of the BMDs. But this sentiment fails to hold up under careful analysis.¹⁰ If and when a voter observes that the BMD-printed ballot is marked with votes that they did not intend, the voter is supposed to alert a poll worker, who is required to void that ballot and allow the voter to mark a fresh ballot. But this situation does not provide usable *evidence* that the BMD was cheating: the voter might be mistaken or lying.¹¹

Therefore, in our hypothetical scenario in which a hacked BMD steals 5% of the votes, and 7% of voters carefully inspect their ballots (and know what to do when they see a mistake), then $7\% \times 5\%$ of voters will alert a pollworker; that is, 1 in every 285 voters will claim their paper ballot was mismarked—if the voters do not assume it was their own error. The BMD would successfully steal “only” 4.65% of the votes.

One might think: “but some voters caught the BMD cheating, red-handed.” But nothing can be done. It is a rare election official who would invalidate an entire election because 1 out of 285 voters complained.¹²

The gap between *voter verifiable* and *voter verified* makes BMDs unacceptable; hacked BMDs can steal the vast majority of the votes they set out to steal, *before* those votes are recorded onto the paper trail. The same analysis applies to a DRE+VVPAT system.

E. All-In-One BMDs

All-in-one BMDs combine the ballot-marking functionality of “pure” BMDs with the scanning/tabulating functionality of optical scanners. In various configurations sold by different manufacturers, the “all-in-one” or “hybrid” BMD may eject the ballot for the voter to inspect before feeding it back into the slot from which it was ejected, or the BMD may display the ballot under glass for voter inspection before retracting it past a scanner.

These machines are even less secure—and less acceptable for use in public elections—than pure BMDs. The same paper path contains both the printer (for marking ballots) and the optical scanner (for scanning ballots). The legitimate software (installed by the manufacturer) presumably will not print additional votes onto the ballot after the voter has inspected it, but hacked software could. The software installed on the BMD has complete control over

¹⁰ See Andrew W. Appel, Richard A. Demillo, & Philip B. Stark, *Ballot-Marking Devices (BMDs) Cannot Assure the Will of the Voters*, ELECTION L. J. (June 16, 2020), <https://doi.org/10.1089/elj.2019.0619> [<https://perma.cc/N3NY-VQ2Y>] (published online).

¹¹ See Matthew Bernhard et al., *Public Evidence from Secret Ballots*, in 10615 ELECTRONIC VOTING, 84, 86 (Robert Krimmer et al. eds., Springer 2017); Tyler Kaczmarek et al., *Dispute Resolution in Accessible Voting Systems: The Design and Use of Audiotegrity*, in 7985 E-VOTING & IDENTITY 127, 131 (James Heather et al. eds., Springer 2013).

¹² See generally Bernhard et al., *supra* note 9.

all the physical functions of the paper path: printing, scanning, and paper transport. Therefore, the hacked computer can print votes on the ballot after the voter's last opportunity to inspect the paper. Even those 7% of voters who carefully inspect their ballots are not safe. The same analysis applies to a DRE+VVPAT system.¹³

F. Internet Voting

Internet voting cannot be secured by any currently known technology.¹⁴ Even if a cryptographic protocol is used to attempt to create an audit trail, the end-user device (phone, computer, or kiosk) is easily hackable. Thus, the voter may indicate a vote for one candidate, but the vote that is encrypted, authenticated, and transmitted may be for another candidate. End-to-end cryptographic paperless voting protocols are an interesting topic for future academic research, but their security and practicality is not mature enough for use in public elections. These scientific facts are well established;¹⁵ we do not discuss them further here.

G. Software Independence, Contestability, Defensibility

By 2004 it was recognized by most experts that paperless DREs were subject to a massive security hole: if fraudulent software was installed in them, that software could steal votes without any way to detect or correct the fraud, or a trustworthy way to recount. In 2008, this understanding was framed in the term “software independence”:

An undetected change or error in its software cannot cause an undetectable change or error in an election outcome.¹⁶

This notion is essential, but still too weak. It very much matters *who* detects the change, and the consequences of this detection. For instance, if an individual voter (amongst the approximately 7% who carefully inspect their BMD-printed ballots) detects an error, that voter has effectively detected a possible error in the election outcome. The election-outcome error is not *undetectable* and the system is software independent.

But in such a case, the election-outcome error is, for all practical purposes, undetectable and uncorrectable by election officials. Voters cannot

¹³ See also Appel et al., *supra* note 10.

¹⁴ NAT'L ACADS. SCIS., ENG'G, & MED., SECURING THE VOTE: PROTECTING AMERICAN DEMOCRACY 92 (2018).

¹⁵ *Id.*

¹⁶ Ronald L. Rivest, *On the Notion of 'Software Independence' in Voting Systems*, 366 PHIL. TRANSACTIONS ROYAL SOC'Y A 1881 (2008).

prove that the votes printed on the paper are not the same as the ones they selected on the BMD. Without any such proof, it would be irresponsible to have a do-over election just on the say-so of a few individual voters.

Appel, DeMillo, and Stark propose the terms “contestable” and “defensible” as more useful in the analysis of voting-system security¹⁷:

A voting system is contestable if, when an undetected change or error in its software causes a change or error in an election outcome, the system can always produce *public* evidence that the outcome is untrustworthy.

A voting system is defensible if, when the reported electoral outcome is correct, it is possible to generate convincing public evidence that the reported electoral outcome is correct—despite any malfunctions, software errors, or software alterations that might have occurred.

A voting system based on BMD-marked ballots is neither contestable nor defensible. A voting system based on hand-marked paper ballots, counted by optical scanners and re-countable (and auditable) by humans, are both contestable and defensible—provided careful procedures are practiced to check administrative processes, physical chain of custody of the ballots, and other physical security measures. Such procedures are called *compliance audits*.

III. RISK-LIMITING AUDITS

If there is a trustworthy paper record of the votes—meaning that a full, accurate hand tabulation of the recorded votes would show the true winners—there is a way to check whether the computers misbehaved: count the votes by hand.

That is an expensive prospect, so some states mandate looking at a sample of ballots instead, i.e., auditing. Generally, statutory audits provide no assurance that, if a reported outcome is wrong, the error will be detected, much less corrected.¹⁸

¹⁷ Appel et al., *supra* note 10.

¹⁸ Some officials claim that the statutory audits check whether the machines are working correctly. But machines never work perfectly. The question is whether they worked well enough, in this election, to find the true winner(s). That is the question a risk-limiting audit answers.

In contrast, a “risk-limiting audit” (RLA) is any post-election procedure that offers the following statistical guarantee¹⁹:

If the reported electoral outcome is wrong, there is a known, pre-determined minimum chance that the procedure will correct the reported outcome.

The maximum chance that the procedure will *not* correct the outcome, if the outcome is wrong, is the “risk limit.” For instance, an RLA with a risk limit of 5% has at least a 95% chance of correcting the reported outcome if the reported outcome is wrong (and no chance of altering a correct reported outcome).

The only possible touchstone for determining the correct outcome and correcting wrong outcomes is the paper trail: an RLA corrects the outcome by conducting a careful, full manual tally of the paper trail. The result of that tally replaces the reported outcome if the two differ.

If the paper trail is trustworthy—i.e., if a full hand tabulation would show who really won—the replacement outcome is the correct electoral outcome, and the overall procedure limits the risk that an incorrect reported outcome will become official. If the paper trail is not trustworthy (for instance, if it has not been kept secure or if it was generated by BMDs), no procedure can limit the risk that an incorrect reported outcome will become official. Indeed, applying an RLA procedure to an untrustworthy paper trail could even replace a correct reported outcome with an incorrect outcome. At best, applying an RLA procedure to an untrustworthy paper trail can check whether tabulation error altered the outcome reflected in the untrustworthy paper trail.

¹⁹ Risk-limiting audits have been endorsed by the National Academies of Sciences Engineering, and Medicine, the Presidential Commission on Election Administration, the American Statistical Association, the League of Women Voters, Common Cause, Verified Voting Foundation, and many other organizations concerned with election integrity. RLAs have been piloted more than fifty times in thirteen U.S. states and in Denmark. They are required by statute in Colorado, Nevada, Rhode Island, and Virginia, and authorized by statute in California and Washington. RLAs were developed in 2007; the first publication is P.B. Stark, *Conservative Statistical Post-Election Audits*, 2 ANN. APPL. STATISTICS 550 (2008). Since then, there have been extensions for other social choice functions (e.g., proportional representation), see P.B. Stark, & V. Teague, *Verifiable European Elections: Risk-limiting Audits for D'Hondt and Its Relatives*, 3 JETS: USENIX J. ELECTION TECH. & SYS. 18 (2014) (for auditing any number of contests simultaneously, for different types of voting equipment, etc.); see also P.B. Stark & M. Lindeman, *A Gentle Introduction to Risk-Limiting Audits*, 10 IEEE SECURITY & PRIVACY 42, (2012) (for a general but still somewhat technical introduction); P.B. Stark, *Sets of Half-Average Nulls Generate Risk-Limiting Audits: SHANGRLA VOTING '20* (forthcoming 2020) [hereinafter Stark, *Half-Average Nulls*] (for the most recent and efficient methods for RLAs).

There are many methods for conducting risk-limiting audits. For instance, a full hand count is a risk-limiting audit, with a risk limit of zero. But, by inspecting randomly selected ballots and using appropriate statistical methods, it is possible to conduct risk-limiting audits much more efficiently—provided the reported electoral outcome is correct.²⁰

IV. COMPLIANCE AUDITS

An RLA procedure that relies on an untrustworthy paper trail, or any audit that purports to ascertain voter intent from an electronic record or from an artifact that the voter did not have the opportunity to check, is “security theater.” There is little reason to believe that a full manual tally of such records would reveal the true winner(s). It is therefore crucial to base audits on voter-verified paper records; to ensure that those records include every validly cast vote exactly once, and no other votes (checking the determination of eligibility, in particular); to ensure that those records remain complete and intact from the moment they are cast through the audit; and to assess the evidence that they are trustworthy. Absent affirmative evidence that the paper trail is a trustworthy record of voter intent—i.e., that tabulating it accurately would show who won according to the intent of every voter who legitimately cast a ballot (in the contests under audit) and no others—the audit might be likely to confirm the incorrect outcome or to change a correct outcome into an incorrect outcome.

The process of assessing the trustworthiness of the paper trail is called a “compliance audit.” Compliance audits should include the following steps, among others:

Ballot Accounting. Check that the number of ballots sent to polling places equals the total number of cast ballots, spoiled ballots, and unvoted ballots. For systems that print ballots on demand, check that the paper stock of voting sheets cast, spoiled, and still blank adds up to the number of sheets sent to the polling place or vote center. Use accountable ballot stock, rather than plain paper, as an important security measure. Check that the number of ballots returned from each polling place does not exceed the number of voters registered at that polling place or the number of pollbook signatures at the polling place. Check that the number of ballots of each style corresponds to the number of ballots of each style reported by the voting system. Ballot counts for this purpose should be based on the physical paper, not on

²⁰ When the reported outcome is incorrect, the audit is *intended* to have a large probability of requiring a full manual tally, so it generally will not save labor in that case.

the voting system: the audit needs external touchstones to check the voting system.

Eligibility. Check signature verification on vote-by-mail ballots. Check the disposition of provisional ballots to ensure that all that were validly cast (and no others) were included in the results. Check that each voter received the correct ballot style based on each voter's eligibility. For vote-by-mail ballots, there should be a record of the ballot style mailed to the voter; for in-person voting, this might require recording (e.g., in pollbooks) the ballot style given to the voter. For provisionally cast ballots, this might be more complicated.

Physical chain of custody. Adopt a formal seal-use protocol²¹ for the tamper-evident seals on ballot boxes and other important records, e.g., use numbered, tamper-evident seals that are hard to forge or bypass; train staff in assessing evidence of tampering; record seal numbers when seals are applied; and check seal numbers against records. Review custody logs. Check that at least two staff members accompanied the ballots whenever ballots were not locked securely or under surveillance. Review surveillance video of the secure ballot storage facility to ensure there was no unauthorized access to ballots.

Due diligence regarding processes, equipment, etc. Review voting equipment event logs. Review any complaints made by voters or anomalies or problems noted by poll workers.

Some of these steps are formally or informally part of the canvass procedure in some jurisdictions. Ideally, the Secretary of State would require these steps (and others) to be conducted in a way that is publicly verifiable and would require jurisdictions to publish the results. Before the election, voter registration databases should be scrutinized and changelogs included. Pre-election "logic and accuracy testing" should include compliance review of the ballot design against EAC usability guidelines²² to ensure that voters will understand the ballot and will not inadvertently overlook some contests or mark ballots incorrectly.

²¹ See generally Andrew W. Appel, *Security Seals on Voting Machines: A Case Study*, 14 ACM TRANSACTIONS INFO. & SYS. SECURITY 2, 1 (2011).

²² See generally U.S. ELECTION ASSISTANCE COMM'N, EFFECTIVE DESIGNS FOR THE ADMINISTRATION OF FEDERAL ELECTIONS (2007), https://www.aiga.org/globalassets/migrated-pdfs/eac_effective_election_design [https://perma.cc/W7UW-S5CQ].

Compliance audits should be a standard part of any recount and not just a precursor to risk-limiting audits. Absent a compliance audit, there is little reason for the public to trust that a recount will find the true winner(s).

V. EFFICIENT RISK-LIMITING AUDITS

The basic strategy behind current methods for risk-limiting audits begins by acknowledging that the reported electoral outcome might be incorrect, then examines randomly selected ballots until either (a) the evidence is convincing that a full manual tally would confirm the reported outcome, or (b) there has been a full manual tally.

There is more than one way to do this. Two basic building blocks are *ballot-polling* and *comparison*. Both can be conducted by randomly selecting either groups of ballots (batch-level audits) or individual ballots (ballot-level audits).²³

Ballot-polling audits are like exit polls, but instead of asking voters how they voted, the audit manually examines randomly selected ballots.²⁴ If a sufficiently large sample of ballots shows a sufficiently large margin in favor of the reported winner, that is evidence that the reported winner really won.²⁵ Ballot-polling audits have the advantage of requiring very little of the voting system: just the reported winners and access to the ballots. They also require local election officials to organize the ballots well enough to draw a random sample of ballots.

Comparison audits compare how the voting system tallied groups of ballots to how humans tally the same physical group of ballots. A group might be, for instance, all ballots tallied in a given precinct or by a given machine, which yields a *batch-level comparison audit*. The most efficient comparison audits use groups consisting of individual ballots, which yield *ballot-level comparison audits*. To conduct a ballot-level comparison audit, the system must report how it interpreted individual ballots in a way that allows the corresponding physical ballot to be identified and retrieved for manual inspection. Such interpretations are called “cast-vote records” or CVRs. The CVR for a ballot lists the voting system’s interpretation of voter intent for each

²³ Ballot-level audits tend to require examining fewer ballots total than audits based on larger batches. Roughly speaking, the number of batches one needs to examine to confirm a contest with a given margin of victory at a given risk limit is about the same, regardless of the batch size. Hence, to attain a given risk limit, an audit that uses batches the size of precincts (say, 500 ballots per batch on average) requires examining about 500 times as many ballots as an audit that uses batches consisting of a single ballot (i.e., a ballot-level audit).

²⁴ Unlike people, ballots always reply, and always reply truthfully, so ballot-polling audits give strong statistical evidence while exit polls generally suffer from large biases.

²⁵ How to quantify the strength of the evidence depends on how the sample is drawn, among other things.

contest on the ballot. Most legacy voting systems cannot report CVRs in a way that the corresponding ballot can be identified and retrieved, but some newer systems have this capability.

One method for conducting a ballot-level comparison audit with a 5% risk limit requires manually inspecting approximately $7/(\text{diluted margin})$ ballots, unless the audit finds errors in the CVRs. The “diluted margin” is the margin of victory in votes, divided by the total number of ballot cards²⁶ in the population from which the sample is drawn (which must include all ballot cards cast in the contest, and may include others). For instance, in the 2018 gubernatorial primary in California, Newsom and Cox advanced to the general election. The margin of Cox over Villaraigosa, the runner-up, was 618,215 votes out of 7,060,646 ballots cast, including undervotes. The diluted margin is thus $618,215/7,060,646 = 8.76\%$. A ballot-level comparison audit with a risk limit of 5% would have required inspecting approximately $7/0.0875 = 80$ ballots selected at random from the entire state (assuming the audit did not find any errors). A ballot-polling audit with a risk limit of 5% would have been expected to examine 443 ballots (assuming that the reported results are correct). For either approach, the amount of work required to justify public confidence in the outcome is *de minimis*.

Most ways of conducting RLAs require a “ballot manifest” describing how ballots are stored. For example, “There are 913 boxes of ballots, numbered 1 through 913. Box 1 contains 301 ballots. Box 2 contains 199 ballots” It is reasonable to require local election officials to construct ballot manifests routinely—if election officials cannot keep track of how much paper there is and where it is, they are not doing their job. Some counties might not currently organize their paper flow in a way that makes constructing ballot manifests possible.

²⁶ A “ballot” often consists of two or more “ballot cards” that contain different contests. Sorting the physical ballot cards into homogeneous groups can greatly reduce the number of cards that must be inspected at random to yield a given number of cards that contain a particular contest.

Ballot manifests should be constructed without relying on the voting system to count the paper; otherwise, we are trusting the voting system to check itself.^{27 28}

VI. RESOURCES FOR RISK-LIMITING AUDITS

Ballot polling requires a ballot manifest and the reported results—the hardware and software requirements are minimal, and open-source code exists for all the computations.²⁹ Batch-level comparison RLAs using precincts as batches generally do not save effort compared to ballot-polling RLAs for typical margins and precinct sizes, but require substantially more “data wrangling.” Ballot-level comparison audits require voting systems that can report cast-vote records for individual ballots in a way that allows the corresponding physical ballot to be retrieved, and vice versa; however, most current voting systems do not have this ability. Ballot-level comparison audits also require exporting those CVRs and “committing” to them in a publicly verifiable way.

RLA methods exist for all common social choice functions used in the U.S., including plurality, vote-for-n plurality (e.g., school boards), super-majority, and instant-runoff voting (IRV, also known as ranked-choice voting, or RCV), as well as proportional representation.³⁰

There is a variety of open-source software to select random samples of ballots and perform risk calculations.³¹ The most difficult aspect of auditing is

²⁷ However, ballot manifests can be augmented by data from the voting system to facilitate audits, provided the audit is designed to take into account the possibility that the voting system data are incorrect. For instance, there are ways to combine cast-vote records with ballot manifests to make it easier to sample ballots that contain specific contests and still ensure that the procedure is an RLA. See Stark, *Half-Average Nulls*, *supra* note **Error! Bookmark not defined.**; Michelle Blom et al., *Sets of Half-Average Nulls Generate Risk-Limiting Audits (SHANGRLA)*, GITHUB, <https://github.com/pbstark/SHANGRLA> [<https://perma.cc/WN2U-FJGU>] (last visited Mar. 17, 2020).

²⁸ Moreover, common human errors include scanning the same box of ballots twice and failing to scan a box of ballots. Scanner mis-picks and errors resulting from clearing scanner paper jams can also cause the number of actual ballots to differ from the number according to the voting system. Relying on the voting system to construct a manifest would miss such errors.

²⁹ See, e.g., P.B. Stark, *Tools for Ballot-Polling Risk-Limiting Election Audits*, UNIV. CAL., BERKLEY: DEP’T OF STATISTICS, <https://www.stat.berkeley.edu/~stark/Vote/ballotPollTools.htm> [<https://perma.cc/DBS2-8LXB>] (last modified Feb. 16, 2017).

³⁰ See Stark, *Half-Average Nulls*, *supra* note 19.

³¹ See, e.g., Blom, *supra* note 27; Stark, *supra* note 29; P.B. Stark, *Tools for Comparison Risk-Limiting Election Audits*, UNIV. OF CAL., BERKLEY: DEP’T OF STATISTICS, <https://www.stat.berkeley.edu/users/stark/Vote/auditTools.htm> [<https://perma.cc/4FUT-8C8X>] (last modified Feb. 26, 2020); pbstark, *auditTools*, GITHUB, <https://github.com/pbstark/auditTools> [<https://perma.cc/V8KH-TN4V>] (last visited Mar. 17,

logistical: coordinating audits of contests that cross jurisdictional lines. That can be facilitated by well-designed software.

In our experience, it takes about two minutes to retrieve a particular randomly selected ballot and transcribe the votes for two or three contests.³² Additional contests take on the order of ten seconds each per audited ballot. The cost of conducting RLAs seems to be very small compared to the overall cost of holding an election. In Colorado, some local election officials report that RLAs are easier than the statutory audits that RLAs replaced, even though the previous audits had little evidentiary value.

A. Audit the Digital Images?

Some vendors are promoting systems that create digital images of ballots. These vendors claim that the images make RLAs easier to perform because fewer (or no) paper ballots need to be inspected. That is incorrect: if a risk-limiting audit relies on images of ballots, it must check that the error in making the images from the voter-verified paper ballots *plus* the error the system made interpreting those images to make cast-vote records is not large enough to cause the electoral outcome to be wrong. It is a mathematical fact that this requires examining at least as many physical ballots as an audit that compares CVRs to a human reading of the paper ballots, without relying on the digital images.³³

VII. PRINCIPLES FOR ELECTION INTEGRITY LEGISLATION

Laws to ensure that election results are trustworthy should satisfy a number of principles:

2020); pbstark, *Tools for SUITE Risk-Limiting Election Audits*, GITHUB, https://github.com/pbstark/CORLA18/blob/master/code/suite_toolkit.ipynb [<https://perma.cc/G6MJ-PA5S>] (last visited Mar. 17, 2020); VotingWorks, *ARLO: Open-Source Risk Limiting Audit Software by VotingWorks*, GITHUB, <https://github.com/votingworks/arlo> [<https://perma.cc/9T8S-PZ99>] (last visited Mar. 17, 2020).

³² The process is much faster if serial numbers are printed on the ballots (after the voted ballot has been dissociated from the voter's identity).

³³ See *supra* note **Error! Bookmark not defined.** for errors that could result in missing or duplicated images. Moreover, there are demonstrations that scanners can inadvertently alter images in ways that would change the appearance of voter intent, including erasing votes. Expecting digital images to accurately reflect voter intent from every validly cast ballot, exactly once, is wishful thinking, even in the absence of hacking. Of course, hacking the scanners or the image processing software is within the technical ability of many undergraduate computer science students.

1. **Require rigorous physical custody of ballots, and compliance audits, as discussed above.** A RLA that relies on an untrustworthy paper record accomplishes little.
2. **Require genuine RLAs.** The procedures and calculations should ensure that whenever an outcome is incorrect, the audit has the requisite chance of leading to a full hand count.³⁴ This entails a number of things:
 - a. **The audit must ascertain voter intent manually—directly from the human-readable marks on the paper ballots the voters had the opportunity to verify.** It is not adequate to rely on digital images of ballots, printout from an electronic record, barcodes, or other artifacts that are not verifiable by the voter or are not tamper evident. Nor is it adequate to re-tabulate the votes electronically, either from images of the ballots or from the original paper. BMD printouts, digital images of ballots, re-printed ballots, and other computer data are not reliable records of voter intent. They can be incomplete, fabricated, or altered (accidentally or maliciously) by software bugs, procedural lapses, or hacking. Statutes should prohibit relying on such things for the determination of voter intent. Making this prohibition explicit is important because, as mentioned above, voting system vendors are marketing technology that purports to facilitate RLAs by allowing auditors to examine digital images of ballots instead of paper ballots. Relying on an electronic record created by the voting system to accurately reflect voter intent amounts to asking a defendant whether the defendant is guilty.
 - b. **The audit must take all validly cast ballots into account.** If ballots are omitted from consideration, for instance vote-by-mail ballots that did not arrive by election night or provisionally cast ballots, the audit cannot be a genuine RLA. Still, there are ways to *begin* an RLA before all ballots are available.
 - c. **The audit must have the ability to correct incorrect outcomes.** This might mean that the audit must take place before results are certified or that the audit can revise already-certified results.

³⁴ The statute should not dictate methods or calculations, only principles. This makes it possible to use improved methods as they are developed or as voting systems are replaced.

3. **Set the risk limit in statute.** Allowing the Secretary of State or local election official to choose the risk limits may create a real or apparent conflict of interest.
4. **Specify how the contests to be audited are selected.**
 - a. **If not every contest will be audited in every election, the selection of contests to audit should involve a random element to ensure that every contest has some chance of being selected.** This ensures that a malicious opponent will not be able to predict whether any particular race will be audited.
 - b. **Every contest not audited with an RLA should be audited using a *risk-measuring audit* instead.**³⁵
 - c. **Statutes must require RLAs on cross-jurisdictional contests—including statewide contests.** Because the point of an RLA is to ensure that reported contest outcomes are correct, every county involved in a particular contest must examine ballots in such a way that the overall cross-jurisdictional procedure is an RLA of that contest. Operationally, auditing cross-jurisdictional contests requires coordination among counties, so each county knows when its portion of the audit can stop. For example, the Secretary of State can tell each jurisdiction how many ballots it needs to draw from each cross-jurisdictional contest in light of the margin and what the audit reveals as it progresses.
5. **The audit sample must not be predictable before the audit starts.** Otherwise, any hacked software would know in which precincts it is safe to cheat. Audits in Colorado, California, Rhode Island, and elsewhere have initialized a random number generator by rolling dice in a public ceremony to ensure that the sample is unknown until that time.³⁶

³⁵ *Risk-measuring* audits are related to *risk-limiting* audits, but they do not have a pre-specified minimum chance of requiring a full manual tabulation when that tabulation would show a different result. In statistical terminology, a risk-measuring audit reports a *P*-value for the hypothesis that a full count would yield a different electoral outcome, based on the audit data. Equivalently, it reports the smallest value for which a risk-limiting audit conducted using that value as its risk limit would have stopped without examining more ballots.

³⁶ Colorado's public ceremony is a good model. See Colorado Secretary of State's Office, *Colorado's risk-limiting audit*, YOUTUBE (Jun. 15, 2018), <https://youtu.be/ysG4pFFmQ-E> [<https://perma.cc/U6GB-VBPP>].

The sample from any collection of ballots should not be selected before election officials have “committed” to the tally of those ballots. For example, nobody should be able to know whether precinct 207 will be audited until the election official has published the tally for precinct 207.³⁷

6. **The public must be able to verify, not merely observe, that the RLA did not stop prematurely.** Among other things, this requires election officials to: disclose the algorithms used to select the sample, calculate the risk and determine when the audit can stop; provide the public the opportunity to observe the selection of the “seed” for drawing the sample; provide adequate public evidence that the paper trail of cast ballots is complete and intact (evidence generated in part by the *compliance audit*); provide the public the opportunity to verify that the correct ballots were inspected during the audit; provide the public the opportunity to observe the voters’ marks on the ballots that were inspected by the audit;³⁸ and in “ballot-level comparison audits,” provide the public proof that the correct cast-vote record was compared to each audited ballot and proof that the full set of cast-vote records yields the reported contest results.

VIII. CONCLUSION

Electronic records of ballots are easy to manipulate by computer hacking. Therefore, voter-verified paper ballots must serve as the auditable evidence that connects the voters’ selections with the election outcome.

Optical scan voting systems, using hand-marked paper ballots designed with usability in mind, have proved to be reliable and highly accurate. These voting systems should be used with compliance-auditable ballot accounting and chain-of-custody procedures, coupled with risk-limiting

³⁷ There are examples (notably, in Cuyahoga County, OH) where election officials altered tallies in precincts selected for recount after the sample was selected to ensure that the inspection would not find any discrepancies. See Kim Zetter, *The Mysterious Case of Ohio's Voting Machines*, WIRED (Mar. 26, 2008, 5:51 PM), <https://www.wired.com/2008/03/the-mysterious/> [<https://perma.cc/2388-JC6G>].

³⁸ It is important to have published rules governing how marks on ballots are to be interpreted in audits and recounts. For instance, if a voter makes a write-in vote for a candidate who is also listed on the ballot, is that a valid vote? If a voter marks a vote for a listed candidate and also writes in that candidate’s name, is that a valid vote? If a voter marks a vote for a candidate, crosses through the mark, and marks a vote for a second candidate, is that a valid vote for the second candidate? If a voter makes a stray mark on the ballot that is distinctive enough to identify the ballot, is the ballot valid?

audits of election tallies, to achieve reliable and trustworthy evidence-based elections.

BMDs were originally envisioned as assistive devices for voters with disabilities who are unable to mark a paper ballot with a pen. Such BMDs have touchscreens, audio interfaces, and ports for other assistive technologies.

Only recently, some states and counties have adopted voting systems that use BMDs for all voters. In light of the insecurity of BMDs—the chasm between *voter-verifiable* and *voter-verified* BMD ballots—hand-marked paper ballots should be the default option presented to all voters, with BMDs available to voters who wish to use them.

Most states already use paper ballots; what we now need to conduct evidence-based elections is better procedures for safeguarding ballots, compliance audits, and risk-limiting audits. These procedures should be enacted in statutes so they have sufficient force of law to truly safeguard our elections against software hacking, insider manipulation, and other threats.

Deploying RLAs (and associated compliance audits) involves the coordination of statistical methods, administrative procedures, paper handling, and more by election administrators across towns, counties, and states. This cannot be done overnight: it requires developing methods appropriate to the election procedures in each state, training officials, educating citizens, practice, and experience. For these reasons, the National Academies of Sciences report recommends that states and local jurisdictions begin with pilot programs and work toward full implementation.³⁹

³⁹ NAT'L ACADS. SCIS., ENG'G, & MED., *supra* note 2.



They May Look and Look, Yet Not See: BMDs Cannot be Tested Adequately

Philip B. Stark¹(✉) and Ran Xie²

¹ University of California, Berkeley, CA, USA
stark@stat.berkeley.edu

² Stanford University, Stanford, CA, USA

Abstract. Bugs, misconfiguration, and malware can cause ballot-marking devices (BMDs) to print incorrect votes. Several approaches to testing BMDs have been proposed. In *logic and accuracy testing* (LAT) and *parallel* or *live* testing, auditors input known test votes into the BMD and check whether the printout matches. *Passive* testing monitors the rate at which voters “spoil” BMD printout, on the theory that if BMDs malfunction, the rate will increase noticeably. We provide lower bounds that show that these approaches cannot reliably detect outcome-altering problems, because: (i) The number of possible voter interactions with BMDs is enormous, so testing interactions uniformly at random is hopeless. (ii) To probe the space of interactions intelligently requires an accurate model of voter behavior, but because the space of interactions is so large, building a sufficiently accurate model requires observing an enormous number of voters in every jurisdiction in every election—more voters than there are in most U.S. jurisdictions. (iii) Even with a perfect model of voter behavior, the required number of tests exceeds the number of voters in most U.S. jurisdictions. (iv) An attacker can target interactions that are intrinsically expensive to test, e.g., because they involve voting slowly; or interactions for which tampering is less likely to be noticed, e.g., because the voter uses the audio interface. (v) Whether BMDs misbehave or not, the distribution of spoiled ballots is unknown and varies by election and possibly by ballot style: historical data do not help much. Hence, there is no way to calibrate a threshold for passive testing, e.g., to guarantee at least a 95% chance of noticing that 5% of the votes were altered, with at most a 5% false alarm rate. (vi) Even if the distribution of spoiled ballots were known to be Poisson, the vast majority of jurisdictions do not have enough voters for passive testing to have a large chance of detecting problems but only a small chance of false alarms.

Keywords: Logic and accuracy testing · Parallel testing · Live testing

1 Introduction

BMDs print votes, often as barcodes or QR codes, together with a human-readable text summary (some BMD printout resembles a hand-marked paper

ballot, HMPB). Jurisdictions including the U.S. state of Georgia, Los Angeles County, California, and Philadelphia, Pennsylvania, recently purchased BMDs for all in-person voters to use.

Bugs, misconfiguration, or malware can make the printed votes and QR codes differ from each other and from the voter's selections. Some have argued that this does not compromise election integrity because voters have the opportunity to inspect BMD printout and to start over if the printout does not match their intended selections; and that since voters can make mistakes hand-marking ballots, HMPBs are no more secure or reliable than BMD printout [19]. We find those arguments unpersuasive:

- In some jurisdictions, the official record of the vote for counts and recounts is the QR code, which voters cannot check.¹
- The arguments equate holding voters responsible for their own errors with holding voters responsible for the overall security of the system [1, 2].
- Most voters *do not* inspect BMD printout [3, 5, 10]. Those who do rarely detect actual errors [3, 13]. To reliably detect errors entails voters taking 3–6 *minutes* to compare a written slate of candidates with the printed selections [12], but voters generally spend less than 3 *seconds* reviewing BMD printout [5, 10].
- If a BMD misprints a voter's selections, only the voter can get evidence of the problem: elections conducted using BMDs are not *contestible* [1].
- If BMDs misbehave, there is no way to determine the correct election outcome because there is no trustworthy paper record of the vote: BMDs are not *strongly software independent* [21].

Concerns about BMDs are not merely hypothetical: BMDs have caused scanners to fail to count votes accurately, to allow voters to vote, and to present all voting options to voters, even after passing LAT [4, 6, 9, 15, 18, 20, 23, 24, 30, 33].

BMD advocates also claim BMDs eliminate ambiguous marks, prevent overvotes, and warn about undervotes (e.g., [19]). But that presumes BMDs function correctly; the rate of truly ambiguous handmade marks is minuscule [1]; and precinct-based optical scanners also protect against undervotes and overvotes (the Voluntary Voting Systems Guidelines, VVSG, require it).² Regardless, elections conducted using BMDs are not trustworthy unless there is a way to ensure that BMD misbehavior did not change any outcome. (If the paper trail itself is not trustworthy, risk-limiting audit procedures do not help because even an accurate full hand count may not reveal who really won.) Elections—and hence BMDs—need to be protected against malicious, technically capable attackers, such as nation states.³ If testing has a high chance of detecting that an outcome

¹ See <https://rules.sos.ga.gov/gac/183-1-15-.03?urlRedirected=yes&data=admin&lookingfor=183-1-15-.03> (last visited 5 May 2022). Audits in Georgia rely on the human-readable text, but legally cannot correct outcomes, and are conducted only for one contest every two years.

² Such protection has been required since VVSG 1.0; see Sect. 2.3.3.2 of [27].

³ The U.S. Senate Intelligence Committee, the Department of Homeland Security, and the FBI concluded that Russian state hackers attacked U.S. elections in 2016 [31].

was altered by a skilled attacker, it also protects against misconfiguration and bugs—which attackers could mimic.

2 Prior Work

Vulnerabilities of particular BMDs are discussed in depth in expert declarations by J. Alex Halderman in *Curling et al. v. Raffensperger et al.*. Theoretical vulnerabilities of various BMD designs are discussed in [1]. [7, 32] discuss testing BMDs; here, we quantitatively investigate their heuristic claims. Three approaches to testing BMDs have been proposed: pre-election logic and accuracy testing (LAT), “live” or “parallel” testing during the election, and “passive” testing by monitoring the spoiled ballot rate. In LAT and parallel testing, auditors make selections on a BMD then check whether the printout accurately reflects those selections. The primary difference is that LAT happens before the election and parallel testing happens during the election. *Passive* testing uses the spoiled ballot rate: if more voters than usual request a do-over, that might be because the machines are malfunctioning.

3 How Much Testing is Enough?

If the paper trail accurately reflects who won, accurate full hand counts and risk-limiting audits (RLAs) can catch and correct wrong outcomes. Here, we study whether testing can establish with high confidence that a paper trail printed by BMDs accurately reflects who won. If not, a recount need not show who really won, and a genuine RLA is impossible.

3.1 Threats and Defenses

We make the following assumptions about BMD threats and defenses:

1. Attackers seek to alter the outcome of one or more contests without being detected. (Some might want to be detected, to undermine public confidence.)
2. Attackers know the testing strategy. This does not preclude the possibility that the strategy will be adaptive or have a random element.
3. Attackers have access to the state history of each BMD, including votes, machine settings, etc.; auditors do not.
4. Attackers have an accurate model of voter behavior in past elections, including political preferences, voting speed, BMD settings, and so on; auditors generally do not, because it would require monitoring voters illegally.
5. Auditors seek to ensure that if any outcome is altered, there is a high chance of detecting it, while keeping the chance of false alarms small.
6. Auditors do not know which contest(s), if any, were altered.
7. Auditors must obey the law and protect voter privacy.

3.2 Jurisdiction Sizes, Contest Sizes, and Margins

U.S. elections are typically administered by counties, townships, or other political units smaller than states. A *ballot style* corresponds to the collection of contests a given voter is eligible to vote in. Typically in the U.S., some contests are on only a fraction of ballot styles in a jurisdiction, in part because many small political units have elections for various offices and measures. Many contests of all sizes are decided by small margins. For instance, in Georgia, U.S., the reported margin in the 2020 presidential election was about 0.2%.

Few votes need to be changed to alter the outcome of small contests and contests with small margins. Conversely, the number of voters in a jurisdiction is an upper bound on the number of passive tests that can be performed and on the sample size to “learn” voter behavior for efficient parallel testing. Thus, jurisdiction size is an important constraint on BMD testing. Since ballot layout, contests, equipment, demographics, political preferences, and other variables vary across and within jurisdictions and malware could affect only some equipment or ballot styles, it is not possible to pool data across jurisdictions to get more power.

Changing votes on 1% of ballots in a jurisdiction can alter the margin of a jurisdiction-wide plurality contest by 2% if there are no undervotes or invalid votes in that contest. If the undervote rate is 30%, then changing votes on 1% of the ballots can change the margin by $0.02/0.7 = 2.9\%$. If a contest is only on 10% of the ballots and the undervote rate in the contest is 30%, altering the votes on 1% of ballots could change the margin in that contest by nearly 29%.

As of 2020, only 1,629 U.S. cities had populations of 100,000 or more, of over 81,363 incorporated places [26]. The 2020 median population of U.S. incorporated areas is 1,201, so about half of the 81,363 incorporated places have turnout of 1,201 or fewer voters. Thus, an attacker does not have to change many votes to alter the outcome of a typical contest for an elected official in a U.S. city or incorporated township. According to [29] the 2020 median turnout in the 6,405 U.S. counties with recorded active voter data was 4,470 voters, and turnout was less than 11,500 voters for more than 2/3 of jurisdictions. In 65.5% of states, more than 50% of counties have fewer than 30,000 active voters. In 85.5% of states, more than 50% of counties have fewer than 100,000 active voters.

3.3 Voting Transactions

We shall call a voter’s interaction with a BMD a *voting transaction* or *transaction* (see Table 1). Transactions are characterized by many variables, including:

- when the transaction starts
- time since the previous voter finished (a measure of polling-place congestion)
- number of transactions before the current transaction
- the voter’s sequence of selections and revisions of selections
- the time to make each selection before taking another action
- whether the voter looks at every page of options in each contest
- the time the voter spends reviewing and revising selections

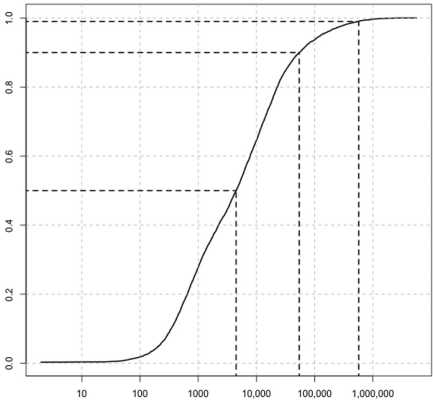


Fig. 1. 2020 turnout by jurisdiction in 3073 counties [29]. Turnout was below 10,000 in $\approx 50\%$ of counties.

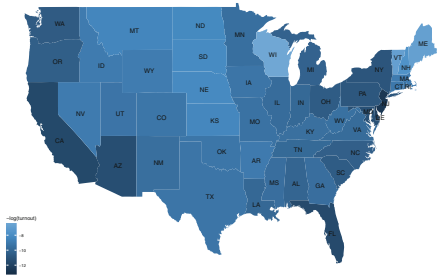


Fig. 2. Median 2020 turnout by jurisdiction in the U.S. [29]

- precisely where voter touches the screen
- BMD settings, including font size, language, use of audio, volume, tempo, pausing, rewinding, use of the sip-and-puff interface, inactivity warnings

Table 1 lists some of the variables and the number of values they can take.⁴ The huge number of possible transactions helps an attacker pick a subset large enough to change an outcome but that auditors are unlikely to probe.

⁴ Table 1 assumes all contests are “vote-for-one.” Ranked-choice voting and multi-winner plurality contests yield more possible transactions. Continuous variables were binned into a few options. VVSG 1.1 [28] requires: (a) Alternative language access is mandated under the Voting Rights Act of 1975, subject to certain thresholds (e.g., if the language group exceeds 5% of the voting age population). (b) The voting system shall provide the voter the opportunity to correct the ballot for either an undervote or overvote before the ballot is cast and counted. (c) An Acc-VS with a color electronic image display shall allow the voter to adjust the color saturation throughout the transaction while preserving the current votes. (d) At a minimum, two alternative display options listed shall be available: 1) black text on white background, 2) white text on black background, 3) yellow text on a black background, or 4) light cyan text on a black background. (e) A voting system that uses an electronic image display shall be capable of showing all information in at least two font sizes. (f) The audio system shall allow the voter to control the volume throughout the voting transaction while preserving the current votes. (g) The volume shall be adjustable from a minimum of 20dB SPL up to a maximum of 100 dB SPL, in increments no greater than 10 dB. (h) The audio system shall allow the voter to control the rate of speech throughout the voting transaction while preserving the current votes. (i) The range of speeds supported shall include 75% to 200% of the nominal rate.

Table 1. Some parameters of BMD transactions and their number of possible values.

Parameter	Optimistic	More realistic
Contests	3	20
Candidates per contest	2	4
Languages	2	13
Time of day	10	20
Number of previous voters	5	140
Undervotes	2 ³	2 ²⁰
Changed selections	2 ³	2 ²⁰
Review	2	2
Time per selection	2	5 ²⁰
Contrast/saturation	—	4
Font Size	2	4
Audio Use	2	2
Audio tempo	—	4
Volume	5	10
Audio pause	—	2 ²⁰
Audio + video	—	2
Inactivity warning	2	2 ²⁰
Total combinations	6.14 × 10 ⁶	3.4 × 10 ⁴⁸

4 Passive Testing

Passive testing sounds an alarm if the number of spoiled ballots exceeds some threshold, t . To ensure that passive testing has a false negative rate (failing to detect altered outcomes) of at most $X\%$, we need to know that the chance that the number of spoiled ballots is greater than or equal to t is at least $X\%$ if BMDs altered any outcome. Conversely, to limit the false alarm rate to at most $Y\%$, we need to know that the chance that the number of spoiled ballots is greater than or equal to t is at most $Y\%$ if BMDs function correctly.

Finding such a value of t is impossible in practice because the distribution of spoiled ballots may depend on ballot design, voting rules, the number of contests, and other things that vary from election to election and place to place—and when BMDs misbehave, also on the number of altered transactions, and the voters and contests affected. Hence, to lower-bound the difficulty, we will assume (optimistically) that the number of spoiled ballots has a Poisson distribution whether BMDs behave correctly or not; but with a rate that depends on the rate of altered transactions. We assume either that 7% of voters will notice errors and spoil their ballots, consistent with the findings of [3], or that 25% of voters will. We consider contest margins of 1%–5% and rates of false positives (false alarms) and false negatives (failing to notice altered outcomes) of 5% and 1%. Results are in Table 2; software to calculate these numbers is in <https://github.com/pbstark/Parallel19>.

Combining Table 2 and Fig. 1 shows that even if the probability distribution of spoiled ballots were known to be Poisson and the spoilage rate when equipment functions correctly were known perfectly, in 2020, in 58.2% of U.S. states fewer than half the counties had enough voters for passive testing to work, even in county-wide contests, on the assumption that 7% of voters whose votes are altered will spoil their ballots.

If turnout is roughly 50%, jurisdiction-wide contests in jurisdictions with fewer than 60,000 voters—22 of California’s 58 counties in 2020 [29]—cannot in

Table 2. Minimum turnout for passive testing with a 5% false negative rate to have at most a 5% false positive rate (cols 3–5) or for passive testing with a 1% false negative rate to have at most a 1% false positive rate (cols 6–8), as a function of the the contest margin (col 1), the percentage of voters who would notice errors (col 2), and the base rate at which voters spoil BMD printout. The number of spoiled ballots is assumed to have a Poisson distribution, with known rate, absent malfunctions. Malfunctions increase the rate by half the margin times the detection rate.

Margin	Voter detection rate	5% error rate			1% error rate		
		Base spoilage rate			Base spoilage rate		
		0.5%	1%	1.5%	0.5%	1%	1.5%
1%	7%	451,411	893,176	1,334,897	908,590	1,792,330	2,675,912
	25%	37,334	71,911	106,627	76,077	145,501	214,845
2%	7%	115,150	225,706	336,160	233,261	454,295	675,242
	25%	9,919	18,667	27,325	20,624	38,039	55,442
3%	7%	52,310	101,382	150,471	106,411	204,651	302,864
	25%	4,651	8,588	12,445	9,870	17,674	25,359
4%	7%	30,000	57,575	85,227	61,385	116,631	171,908
	25%	2,788	4,960	7,144	5,971	10,312	14,681
5%	7%	19,573	37,245	54,932	40,156	75,671	110,989
	25%	1,838	3,274	4,689	4,036	6,849	9,650

principle limit the chances of false positives and false negatives to 5% for margins below 4%, even under these optimistic assumptions. For contests that involve only part of a jurisdiction, the situation is worse.

4.1 Targeting Vulnerable Voters

The analysis above assumes that all voters are equally likely to detect errors and spoil their ballots. But an attacker can use BMD settings, state history, and session data to target voters who are less likely to notice problems.

Voters with Visual Impairments. Approximately 0.8% of the U.S. population is legally blind; approximately 2% age 16 to 64 have a visual impairment [16]. Current BMDs do not provide voters with visual impairments a way to check the printout. If an attacker only alters votes when the voter uses the audio interface or large fonts, detection may be very unlikely.

Voters with Motor Impairments. Some BMDs allow voters to print and cast a ballot without looking at it, for instance the ES&S ExpressVote[®] with “Auto-cast,” aka “permission to cheat” [1]. The attacker can change every vote cast using Autocast, with zero chance of detection.

Voters who use Languages other than English. U.S. law requires some jurisdictions to provide ballots in languages other than English. For instance, Los Angeles County, CA, provides voting materials in 13 languages [14]. In 2013, roughly 26% of voters in Los Angeles County spoke a language other than English at home [14]. It is our understanding that BMDs generally print only in English. If voters who use a foreign language on the BMD are unlikely to check the English-language printout, an attacker could change the outcome of contests with large margins with little chance of detection.

Fast and Slow Voters. An attacker can monitor how long it takes voters to make their selections, whether they change selections, how long they review the summary screen, etc. A voter who spends little time reviewing selections onscreen may be unlikely to review the printout carefully. Conversely, a voter who takes a very long time to make selections or changes selections repeatedly might find voting difficult or confusing and be unlikely to notice errors.

Thus, it is in the attacker's interest to target the same groups of voters BMDs are supposed to help: voters with visual impairments, voters with limited dexterity, voters who use a language other than English, and voters with cognitive disabilities.

4.2 FUD Attacks on Passive testing

Even under ideal circumstances, passive testing does not produce direct evidence of problems; it does not identify which ballots or contests have errors; and it does not provide any evidence about whether problems changed outcomes. Relying on spoiled ballots as a sign of fraud opens the door to a simple, legal way to undermine elections: encourage voters to spoil ballots.

5 LAT and Parallel Testing

Suppose that malware alters one or more votes with probability p , independently across transactions, uniformly across voters—regardless of the voter's selections or any aspect of the transaction that the attacker can ascertain. Then if auditors make n tests, the chance that the BMD will alter at least one of the votes in at least one of the tests—and the attack will be detected—is $1 - (1 - p)^n$. For $p = 0.01$, $n = 300$ tests would give a 95% chance of detecting a problem.

A BMD can handle roughly 140 transactions per day. Testing enough to have a 95% chance of detecting a 1% problem on one BMD would leave no time for voters to use that BMD. Even for pre-election LAT, where capacity for actual voters is not an issue, conducting 300 “typical” tests would take about 25 h.

If there were a large number of machines known to have been (mis)programmed identically, tests could be spread across them. But there are many small contests that need to be tested in conjunction with all other contests that appear on any ballot style that contains them, and there is no guarantee that all BMDs in a jurisdiction are programmed identically.

This threat model is completely unrealistic. An attacker who wants Alice to beat Bob will not alter votes for Alice: it would needlessly increase the chance of detection. And as discussed in Sect. 4.1, rather than randomly changing votes for Bob into votes for Alice, an attacker can target transactions that auditors are unlikely to probe.

Setting aside specific machines for testing facilitates a “Dieselgate” type attack [11], as does conducting tests on a schedule, as suggested by [7]. Tests need to be unpredictable—with respect to the specific BMDs tested, time, vote pattern, duration, and other characteristics of voting transactions—or attackers

can avoid detection by altering only transactions that do not correspond to any test. There may be pressure to reduce testing when BMDs are busy, to reduce waiting times. Because malware can monitor the pace of voting, reducing testing when machines are busy makes it easier to avoid detection.

An attacker need not alter many transactions to change the outcome of small contests and contests with small margins. The fewer votes altered, the more tests required to ensure a large chance of detection. To test efficiently, tests should sample more common transactions with higher probability. Attackers might be able to estimate of the distribution of transactions using malware installed on BMDs in previous elections, but testers will not, since it involves tracking voter behavior at a level of detail that violates voter privacy. See assumptions 4 and 7 and Sect. 5.2.

Auditors do not know which contest(s) and candidate(s) are affected. To have a large chance of detecting interference, there needs to be a large chance of testing a transaction the attacker alters. Attackers can target transactions that are intrinsically expensive to test, e.g., transactions that take longer than 10 min, transactions in which the voter changes some number of selections, transactions that display the ballot in a language other than English, transactions that use the audio interface at a reduced tempo, etc.

5.1 Lower Bounds on the Difficulty of Parallel Testing

We now study an idealized version of parallel testing, where auditors can tell whether a random sample of BMD printouts accurately show the voters' selections. Suppose a contest has 4,470 voters, the median jurisdiction turnout in 2020. Suppose that malware alters votes in 23 transactions, which could change a margin by more than 1% in a jurisdiction-wide contest. How many randomly selected printouts would need to be checked to have at least a 95% chance of finding at least one with an error? The answer is the smallest n such that

$$\frac{4470 - 23}{4470} \cdot \frac{4469 - 23}{4469} \cdots \frac{4470 - (n - 1) - 23}{4470 - (n - 1)} \leq 0.05, \quad (1)$$

i.e., $n = 546$ printouts, about 12.2% of the transactions, corresponding to testing each BMD several times per hour.

Conversely, suppose auditors randomly check 13 printouts per day per machine (on average, testing hourly for a 13-hour day, $\approx 9.2\%$ of BMD capacity). To have at least a 95% chance of detecting that the outcome of a contest with a 1% margin was altered, there would need to be at least 6,580 voters in the contest (almost 150% the median turnout in jurisdictions across the U.S.), corresponding to 47 BMDs, even under these unrealistically optimistic assumptions.

5.2 Building a Model of Voter behavior

In practice, auditors cannot check whether voters' BMD printout is correct. Instead of sampling voters' actual transactions in the election, they will have

to come up with their own test transactions. Testing transactions uniformly at random from all possible transactions is doomed because the number of possible transactions is so large. To mimic voters, auditors might consider sampling from P , the population distribution of voting transactions, i.e., the fraction of voters who use the BMD in each of the $S = 6.14 \times 10^6$ ways in the optimistic estimate in Table 1. Suppose an attacker wants to change the outcome of a contest with a margin of m , expressed as a fraction of ballots cast (rather than as a number of votes). The attacker only needs to change a fraction $m/2$ of the transactions to change the margin by m . To have probability at least $1 - \alpha$ of detecting a change to the outcome of any contest with true margin m , auditors must test in a way that has probability at least $1 - \alpha$ of sampling at least once from every subset of transactions that contains a fraction $m/2$ of the transactions. If auditors could sample transactions independently at random from P , each sample transaction would have probability $1 - m/2$ of *not* being one of the altered transactions. The chance that t randomly selected transactions would not include one that is altered would be $(1 - m/2)^t$. Thus the number of transactions auditors would need to test is the smallest t for which

$$(1 - m/2)^t \leq \alpha, \text{ i.e.,} \tag{2}$$

$$t \geq \frac{\log \alpha}{\log(1 - m/2)}. \tag{3}$$

This is essentially Eq. 1 for sampling with replacement; the two are indistinguishable when t is small compared to the total number of possible transactions. A key difference is that in Eq. 1, auditors are sampling from the actual transactions in the election, while in Eq. 3, auditors are sampling from a model, the frequency distribution of of transactions.

In practice, the auditors do not know P —they will have to estimate it by monitoring voters. In reality, this is impossible to do well: (i) In a given election, P will depend on the particular contests on the ballot and the particular voters who participate, both of which change from election to election. (ii) The variables that characterize a voting transaction include the voter’s selections and details about how the voter uses the BMD, so collecting the data would violate voter privacy illegally. To get a sense of the *statistical* difficulty of the problem, we ignore these practical difficulties. If auditors could select voters at random (with replacement) and observe in detail how they use the BMD—all the variables in Table 1—that would yield independent, identically distributed (IID) draws from P , which could be used to make an estimate, \hat{P} . If \hat{P} differs too much from P , no number of tests will suffice, because \hat{P} might estimate that the frequency of a transaction is zero when in fact it is sufficiently frequent that altering it could change an outcome. (By assumption 4, above, the attacker knows P and hence can exploit differences between \hat{P} and P .) How many voters would auditors have to observed to ensure (with sufficiently high probability) that \hat{P} is accurate enough for parallel testing?

The L_1 distance between two distributions bounds the difference in the probability they assign to any set ($|\hat{P}(A) - P(A)| \leq \|\hat{P} - P\|_1/2$). If $\|\hat{P} - P\|_1 \geq m$,

there may be a set A of transactions for which $P(A) = m/2$ but $\hat{P}(A) = 0$, so changing votes for transactions in A could alter some margin⁵ by m , with zero chance of detection, no matter how many tests are performed, if the tests are drawn from \hat{P} rather than P .

We cannot guarantee that $\|\hat{P} - P\|_1 \leq \varepsilon$ with *certainty*, but by observing enough randomly selected voters, we can ensure that the chance that $\|\hat{P} - P\|_1 > \varepsilon$ is at most β . If $\alpha \leq \beta$, even an infinite number of tests drawn from \hat{P} may not suffice to guarantee chance at least $1 - \alpha$ of detecting outcome-changing manipulations. If $\alpha > \beta$, to guarantee chance at least $1 - \alpha$ of catching an outcome-changing error, the minimum number of tests required is

$$\min \left\{ t : (1 + \varepsilon/2 - m/2)^t \leq \frac{\alpha - \beta}{1 - \beta} \right\}. \quad (4)$$

Minimax Lower Bounds. Suppose auditors draw an IID sample of n transactions from P , a frequency distribution on S possible transactions. Let \mathcal{M}_S denote the collection of all frequency distributions for those transactions. Then the training sample size n must be at least large enough to ensure that the L_1 error of the best estimator \hat{P} is unlikely to exceed ε , provided $P \in \mathcal{M}_S$:

$$\inf_{\hat{P}} \sup_{P \in \mathcal{M}_S} \Pr\{\|\hat{P} - P\|_1 \leq \varepsilon\} \geq 1 - \beta. \quad (5)$$

Theorem. ([8]) For any $\zeta \in (0, 1]$,

$$\begin{aligned} \inf_{\hat{P}} \sup_{P \in \mathcal{M}_S} \mathbb{E}_P \|\hat{P} - P\|_1 &\geq \frac{1}{8} \sqrt{\frac{eS}{(1 + \zeta)n}} \mathbb{1} \left(\frac{(1 + \zeta)n}{S} > \frac{e}{16} \right) \\ &\quad + \exp \left(-\frac{2(1 + \zeta)n}{S} \right) \mathbb{1} \left(\frac{(1 + \zeta)n}{S} \leq \frac{e}{16} \right) \\ &\quad - \exp \left(-\frac{\zeta^2 n}{24} \right) - 12 \exp \left(-\frac{\zeta^2 S}{32(\ln S)^2} \right), \end{aligned} \quad (6)$$

where the infimum is over all \mathcal{M}_S -measurable estimators \hat{P} .

Lemma. Let X be a random variable with variance $\text{Var} X \leq 1$, and let $\beta \in (0, 1)$. If $\Pr\{X \geq \mathbb{E}X + \lambda\} \leq \beta$ then $\lambda \geq -\sqrt{\beta/(1 - \beta)}$.

Proof. Suppose $\lambda \geq 0$. Then $\lambda \geq -\sqrt{\beta/(1 - \beta)}$. Suppose $\lambda < 0$. By Cantelli's inequality and the premise of the lemma,

$$\beta \geq \Pr\{X \geq \mathbb{E}X + \lambda\} \geq 1 - \frac{\sigma^2}{\sigma^2 + \lambda^2} = \frac{\lambda^2}{\sigma^2 + \lambda^2} \geq \frac{\lambda^2}{1 + \lambda^2}. \quad (7)$$

⁵ The set of undetectable shifts of $m/2$ votes might not include the one that any particular attacker seeks; this bound is worst-case across hypothetical attackers and distributions of transactions.

Solving for λ yields the desired inequality. \square .

Now $0 \leq \|\hat{P} - P\|_1 \leq 2$, so $\text{Var}\|\hat{P} - P\|_1 \leq 1$. By the lemma, we need $\lambda \geq -\sqrt{\beta/(1-\beta)}$ to ensure that $\Pr\{\|\hat{P} - P\|_1 \geq \mathbb{E}\|\hat{P} - P\|_1 + \lambda\} \leq \beta$. If $\|\hat{P} - P\|_1 \geq 2r$, there can be a set of transactions τ such that $P(\tau) = m/2$ but $\hat{P}(\tau) = 0$, so if tests are generated randomly according to \hat{P} there is zero probability of testing any transaction in τ , no matter how many tests are performed. Thus if $\Pr\{\|\hat{P} - P\|_1 \geq m\} > \alpha$, even an infinite number of tests cannot guarantee chance at least $1 - \alpha$ detecting that a fraction $m/2$ of the transactions were altered, enough to wipe out a margin of m . By the lemma, that is the case if $m < \mathbb{E}\|\hat{P} - P\|_1 - \sqrt{\alpha/(1-\alpha)}$, i.e., if $\mathbb{E}\|\hat{P} - P\|_1 > m + \sqrt{\alpha/(1-\alpha)}$.

The theorem gives a family of lower bounds on $\mathbb{E}\|\hat{P} - P\|_1$ in terms of n . If the lower bound exceeds $m + \sqrt{\alpha/(1-\alpha)}$, testing by drawing transactions from \hat{P} cannot protect against all outcome-changing errors. The bound grows with S , the number of possible transactions. To be optimistic, we use an unrealistically small value $S = 6.14 \times 10^6$ (Table 3).⁶

To guarantee a 95% chance of detecting that $m/2 = 5\%$ of transactions were altered, which could change jurisdiction-wide margins by 10% or more, the training sample would need to include at least 1.082 million transactions, even if auditors could conduct an infinite number of parallel tests. That is larger than the turnout in 99.7% of U.S. jurisdictions in 2020 [29]; it is roughly 0.5% of the U.S. voting population. To guarantee 99% chance of detecting that 0.5% of transactions were altered, which could change jurisdiction-wide margins by 1% or more, would require observing 3.876 million voters in complete, privacy-eliminating detail—more than the turnout in 99.9% of U.S. jurisdictions in 2020 [29], roughly 1.9% of the U.S. voting population.

Table 3. Lower bound on the sample size (col 4) required to estimate the distribution of voting transactions well enough to ensure the probability (col 1) of detecting the manipulation of the fraction of transactions (col 3) using some number of tests (col 2), if the support of the distribution of transactions has $S = 6.14 \times 10^6$ points.

Confidence level	Maximum tests	Altered votes	Bound (millions)
99%	2000	0.5%	3.87
		1%	3.58
		3%	2.69
		5%	2.09
95%	2000	0.5%	1.67
		1%	1.59
		3%	1.31
		5%	1.10
99%	Inf	0.5%	3.73
		1%	3.46
		3%	2.61
		5%	2.04
95%	Inf	0.5%	1.65
		1%	1.57
		3%	1.29
		5%	1.08

⁶ Software implementing the calculations is in <https://github.com/pbstark/Parallel19>.

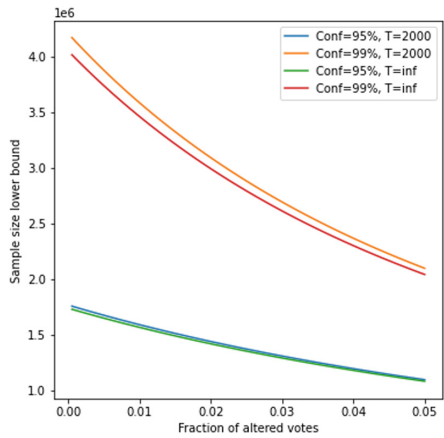


Fig. 3. Minimum training sample sizes as a function of the fraction of altered votes.

6 Complications

Reality is worse than the optimistic assumptions in our analyses:

Margins are not Known in Advance. Margins are not known until the election is over, when it is too late to do more testing if contests have narrower margins than anticipated. Testing to any pre-defined threshold, e.g., a 95% chance of detecting changes to 0.5% of the votes in any contest, will not always suffice.

Tests have Uncertainty. If the BMD printout reflects the wrong electoral outcome, a perfect full manual tally, recount, or risk-limiting audit based on BMD printout will confirm that wrong outcome. Suppose one could design practical parallel tests that had a 95% chance of sounding an alarm if BMDs alter $\geq 0.5\%$ of the votes in any contest. A reported margin of 1% or less in a plurality contest is below the “limit of detection” of such tests. Would laws require a runoff whenever a reported margin is below the limit of detection of the tests?

Special Risks for Some Voters. As discussed above in Sect. 4.1, BMDs can be used to selectively disenfranchise voters with disabilities and voters whose preferred language is not English. Indeed, the attacker’s best strategy is to target such voters, in part because poll workers might more likely to think that complaints by such voters reflect voter mistakes rather than BMD malfunctions.

The only Remedy is a New Election. If a BMD is caught misbehaving, it should be removed from service and all BMDs in that jurisdiction should be investigated. But there is no way to determine the correct outcome or which votes were affected: BMDs are not *strongly software independent* [21].

7 Conclusion

We show that to protect against outcome-altering BMD malfunctions requires orders of magnitude more testing than is feasible. To our knowledge, no jurisdiction has conducted *any* parallel testing of BMDs of the kind suggested by [7, 32], much less enough to reliably detect outcome-changing errors, bugs, or hacks.

Even if it were possible to test enough to get high confidence that no more than some threshold percentage the votes were changed in any contest, fairness would demand a runoff in contests decided by less than that threshold.

Some BMDs may be the best extant technology for voters with particular disabilities to mark and cast a paper ballot independently. But many BMDs are poorly designed. Some have easily exploited security flaws [1] and some do not enable voters with common disabilities to vote independently [22, pp. 68–90]. To our knowledge, no VVSG-certified BMD system provides a means for blind voters to check whether the printout matches their intended selections.

Using BMDs makes elections less trustworthy, less resilient, less transparent, more fragile, and more expensive [1, 17]. BMDs have failure modes that hand-marked paper ballots do not have, and lack resilience when failures occur [1]. BMDs shift the burden of ensuring that voting equipment functions correctly from officials to voters, but do not provide voters any way to prove that they observed problems, if they do; nor can election officials show that outcomes are correct despite any problems that might have occurred [1]. BMDs undermine the ability of election officials to provide affirmative evidence that outcomes are correct, the fundamental principle of “evidence-based elections” [2, 25].

Voters who use BMDs should be urged to bring a written list of their selections to the polls to check against BMD printout, and to request a fresh ballot if the printout does not match their intended selections. Election officials should track spoiled BMD printouts. There should be research on how to encourage voters to check BMD printout and report discrepancies, how to ensure the checks are accurate, and how to ensure that any reported problems are accountably and transparently recorded, addressed, and publicized; these issues also arise in end-to-end cryptographically verifiable (E2E-V) voting systems. For the foreseeable future, prudent election administration requires keeping the use of BMDs to a minimum.

Acknowledgements. We are grateful to Yanjun Han and Tsachy Weissman for helpful conversations about minimax L_1 estimation, and to Peter Rønne and anonymous referees for helpful comments and suggestions.

References

1. Appel, A., DeMillo, R., Stark, P.: Ballot-marking devices (BMDs) cannot assure the will of the voters. *Election Law J.* **19**, 432–450 (2020). <https://doi.org/10.1089/elj.2019.0619>
2. Appel, A., Stark, P.: Evidence-based elections: create a meaningful paper trail, then audit. *Georgetown Law Technol. J.* **4**(2), 523–541 (2020). <https://georgetownlawtechreview.org/wp-content/uploads/2020/07/4.2-p523-541-Appel-Stark.pdf>
3. Bernhard, M., et al.: Can voters detect malicious manipulation of ballot marking devices? In: 41st IEEE Symposium on Security and Privacy, pp. 679–694. IEEE (2020). <https://doi.org/10.1109/SP40000.2020.00118>
4. Cillizza, C.: How did Georgia get it so wrong (again)? (2020). <https://www.cnn.com/2020/06/10/politics/georgia-primary-vote-brian-kemp/index.html>
5. DeMillo, R., Kadel, R., Marks, M.: What voters are asked to verify affects ballot verification: a quantitative analysis of voters’ memories of their ballots (2018). <https://doi.org/10.2139/ssrn.3292208>
6. Fowler, S.: State outlines fix for error that halted election testing. Georgia Public Broadcasting (2020). <https://www.gpb.org/news/2020/09/29/state-outlines-fix-for-error-halted-election-testing>
7. Gilbert, J.: Ballot marking verification protocol. <http://www.juangilbert.com/BallotMarkingVerificationProtocol.pdf> (2019)
8. Han, Y., Jiao, J., Weissman, T.: Minimax estimation of discrete distributions. In: 2015 IEEE International Symposium on Information Theory (ISIT), pp. 2291–2295. IEEE (2015)
9. Harte, J.: Exclusive: Philadelphia’s new voting machines under scrutiny in Tuesday’s elections. Reuters (2020). <https://in.reuters.com/article/usa-election-pennsylvania-machines/exclusive-philadelphias-new-voting-machines-under-scrutiny-in-tuesdays-elections-idINKBN2382D2>
10. Haynes, A., III, M.H.: Georgia voter verification study. <https://s3.documentcloud.org/documents/21017815/gvvs-report-11.pdf> (2021). Accessed 31 Oct 2021
11. Hotten, R.: Volkswagen: the scandal explained. <https://www.bbc.com/news/business-34324772> (2015)
12. Kortum, P., Byrne, M., Azubike, C., Roty, L.: Can voters detect errors on their printed ballots? Absolutely. <https://arxiv.org/abs/2204.09780> (2022)
13. Kortum, P., Byrne, M., Whitmore, J.: Voter verification of ballot marking device ballots is a two-part question: can they? mostly, they can. do they? mostly, they don’t. *Election Law J. Rules Polit. Policy* 243–253 (2021). <https://doi.org/10.1089/elj.2020.0632>
14. Los Angeles county clerk: multilingual services program (2020). <https://www.lavote.net/home/voting-elections/voter-education/multilingual-services-program/multilingual-services-program>
15. Mehrotra, K., Newkirk, M.: Expensive, glitchy voting machines expose 2020 hacking risks (2019). <https://www.bloomberg.com/news/articles/2019-11-08/expensive-glitchy-voting-machines-expose-2020-hacking-risks>
16. National federation of the blind: blind statistics (2019). <https://www.nfb.org/resources/blindness-statistics>
17. Perez, E., Miller, G.: Georgia state election technology acquisition: a reality check. https://trustthetvote.org/wp-content/uploads/2019/03/06Mar19-OSETBriefing_GeorgiaSystemsCostAnalysis.pdf (2019)

18. Previti, E.: Northampton officials unanimously vote ‘no confidence’ in ExpressVote XL voting machine (2019). <https://papist.org/2019/12/20/northampton-officials-unanimously-vote-no-confidence-in-expressvote-xl-voting-machine/>
19. Quesenbery, W.: Why not just use pens to mark a ballot? (2018). <https://civicedesign.org/why-not-just-use-pens-to-mark-a-ballot/>
20. Riggall, H.: Up to 157 incorrect ballots cast on first day of early voting, Cobb elections director says. Marietta Daily Journal, Ga (2022). <https://news.yahoo.com/157-incorrect-ballots-cast-first-092000088.html>
21. Rivest, R.: On the notion of ‘software independence’ in voting systems. Phil. Trans. R. Soc. A **366**(1881), 3759–3767 (2008)
22. Secretary of the commonwealth of Pennsylvania: report concerning the examination results of election systems and software EVS 6012 with DS200 precinct scanner, DS450 and DS850 central scanners, ExpressVote HW 2.1 marker and tabulator, ExpressVote XL tabulator and electionware EMS. <https://www.dos.pa.gov/VotingElections/Documents/Voting%20Systems/ESS%20EVS%206021/EVS%206021%20Secretary%27s%20Report%20Signed%20Including%20Attachments.pdf> (2018)
23. Shortell, T., Tatu, C.: Here’s why northampton county’s voting machines went wrong, county executive says. The Morning Call 12 December 2019 (2019)
24. Sneed, T.: Will L.A.’s voting overhaul be an industry disrupter or the next election debacle? (2020). <https://talkingpointsmemo.com/news/will-l-a-s-voting-overhaul-be-an-industry-disrupter-or-the-next-election-debacle>
25. Stark, P., Wagner, D.: Evidence-based elections. IEEE Secur. Priv. **10**, 33–41 (2012)
26. U.S. Census Bureau: city and town population totals: 2010–2020 (2020). <https://www2.census.gov/programs-surveys/popest/datasets/2010-2020/cities/>
27. U.S. Election Assistance Commission: voluntary voting systems guidelines 1.0, December 2005. https://www.eac.gov/sites/default/files/eac_assets/1/28/VVSG.1.0.Volume.1.PDF
28. U.S. Election Assistance Commission: voluntary voting system guidelines version 1.1. U.S. Election Assistance Commission (2015). https://www.eac.gov/sites/default/files/eac_assets/1/28/VVSG.1.1.VOL.1.FINAL1.pdf
29. U.S. Election Assistance Commission: The U.S. election assistance commission’s 2020 election administration and voting survey (EAVS) (2020). <https://eavsportal.com/>
30. U.S. Election Assistance Commission: report of investigation, dominion voting systems D-suite 5.5-B. Williamson County, Tennessee. https://www.eac.gov/sites/default/files/TestingCertification/EAC_Report_of_Investigation_Dominion_DSuite_5.5.B.pdf (2022)
31. U.S. Senate Intelligence Committee: Russian targeting of election infrastructure during the 2016 election: summary of initial findings and recommendations. <https://www.burr.senate.gov/imo/media/doc/RussRptInstlmt1-%20ElecSec%20Findings,Recs2.pdf> (2018). Accessed 3 June 2020
32. Wallach, D.: On the security of ballot marking devices. Ohio State Technol. Law J. **16**(2), 558–586 (2020)
33. Zetter, K.: Los Angeles County’s risky voting experiment (2020). <https://www.politico.com/news/2020/03/03/los-angeles-county-voting-experiment-119157>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



There is no Reliable Way to Detect Hacked Ballot-Marking Devices

Philip B. Stark

University of California, Berkeley

21 August 2019

Abstract. Election system vendors are marketing ballot-marking devices (BMDs) as a universal system, and some states are deploying them for all voters, not just those who need a BMD to vote independently. Like all devices with CPUs, BMDs can be hacked, misprogrammed, or misconfigured. BMD printout might not reflect what the BMD screen or audio confirmed. If a voter complains that the BMD altered votes, officials have no way to tell whether there was a BMD malfunction, the voter erred, or the voter is attempting to cast doubt on the election. Pre-election logic and accuracy (L&A) tests have little ability to detect outcome-changing problems, in part because BMDs know the time and date, and in part because it is in practice impossible to simulate the full range of voter interactions with the device. *Parallel* or *live* tests of BMDs on election day address the first problem but not the second. In practice, neither L&A nor parallel tests can probe all combinations of voter preferences, device settings, ballot language, duration of voter interaction, input and output interfaces, and other variables that could include a enough votes to change election outcomes. Even if less parallel testing sufficed, it would still require extra BMDs, new infrastructure for creating test patterns, new infrastructure for reporting test results, additional polling-place staff, and additional staff training. And if parallel testing discovers an error, the only remedy is to hold a new election: there is no way to reconstruct the correct election result from an untrustworthy paper trail. Minimizing the number of votes cast using BMDs is prudent election administration.

Keywords: elections, logic and accuracy testing, parallel testing, live testing

In theory there's no difference between theory and practice, but in practice there is.

— Jan L.A. van de Snepscheut

1 Introduction

Voting system vendors are promoting the use of ballot-marking devices (BMDs) for all voters, and several states are currently contemplating purchasing enough BMDs for all voters. While this option is more profitable for vendors, deploying BMDs for all voters is more expensive and less reliable than using BMDs as needed for accessibility (Perez 2019) and letting voters who are able to hand mark a ballot mark their ballots by hand.

It is commonly argued that security problems in BMDs are not an issue for election integrity because voters have the opportunity to inspect the BMD printout before casting it, and to spoil the printout and start over if the printout does not match their intended selections. It is also commonly argued that since voters can make mistakes marking a

ballot by hand, hand-marked ballots are no more secure or reliable than BMD printout. Both arguments are myopic: they do not take the overall security of the election into account, and confuse holding a voter responsible for the voter’s own errors with holding voters responsible for the overall security of a vulnerable electronic system.

Current BMDs and the protocols for their use make voters responsible for testing BMD security, but do not give them the tools they need to do it. In particular, (Stark 2019) and (Appel, DeMillo, and Stark 2019) point out an intrinsic security gap in current ballot-marking devices: if what the device prints on the paper differs from what the voter expressed to the device or confirmed on a review screen, only the voter knows for sure. The system does not generate any evidence the voter can present to a pollworker or election official to prove there was a discrepancy.¹ There is no way for an election official to tell whether a complaint that a BMD altered selection indicates a machine error, a voter error, or a cry of “wolf” intended to cast doubt on election results.

As a result, error or malfeasance could change a large percentage of votes without inducing election officials to act, beyond possibly offering voters who notice errors another chance to mark a ballot. Worse, if a problem is detected, the only remedy is to hold a new election: there is no way to figure out which printed votes were affected, nor what should have been printed.

Pre-election “logic and accuracy” (L&A) testing should be a routine precaution, but it is not well suited to discovering malware or subtle bugs. It is not generally possible to probe many of the possible vote combinations during L&A testing: to do so would take far too long. There are a number of ways a BMD could “know” that it is being tested, for instance, if the test is not during polling hours on election day. Even if one resets the internal clock on a BMD, the way L&A testers interact with the device is unlikely to closely match how voters interact with the device. The difference can be used as “cryptic knock” like that at the root of the VW “dieselgate” scandal: the device could behave one way when it is being tested, and differently when used by actual voters.

In to address some of the shortcomings of L&A testing, some computer scientists have proposed using “parallel testing” or “live testing.” In parallel testing, pollworkers or trusted agents (*testers*) use a BMD to mark ballots on election day, but do not cast those ballots. Their input selections are compared to what the BMDs printed. Even a single discrepancy is evidence of a serious problem: that BMD alters votes, and others might, too.

Because BMDs within a jurisdiction are expected to be running the same software and to be configured identically, a discrepancy on one machine casts doubt on an entire election.²

How much testing is required? Juan Gilbert suggests testing once per hour.³ Dan Wallach suggests testing approximately at random, with no detail about how and when to test (Wallach 2019). While “enough” suitably designed random testing can provide assurance,

¹ Voters could take “ballot selfie” videos of the interaction with the BMD, the review screen, and the printed output, but that would compromise voter privacy and is illegal in many jurisdictions. Moreover, it would not be hard for malicious voters to fake problems by revising their selections before printing, or by constructing a “deep fake.” That opens elections to attacks on public confidence by spreading unwarranted fear, uncertainty, and doubt (FUD).

² However, BMDs cannot be *assumed* to be programmed and configured identically: if one BMD behaved correctly, that does not mean every BMD behaved correctly. An attacker might hack only some of the machines, and hacking might be designed to affect different machines differently. Indeed, examples below show how that can help an attacker avoid detection.

³ <http://www.juangilbert.com/BallotMarkingVerificationProtocol.pdf>, last visited 15 August 2019.

“enough” turns out to be too much to be practical, both for parallel testing and for L&A testing.⁴

2 A Naive Calculation

Gilbert and Wallach seem to have the following calculation in mind. (Indeed, the calculations in section 5.1 of (Wallach 2019) are exactly of this form.)

Suppose that a BMD alters one or more votes on a ballot with probability p , independently across ballots—regardless of the selections made on the BMD, the time of day, or any of the BMD settings (e.g., whether the audio interface is being used, the language the ballot is displayed in, the font size, etc.) and without regard for aspects of voter behavior that the BMD can monitor (e.g., how long the voter takes to make selections, whether the voter changes any selections, whether the voter backs up to previous pages, etc.). Then if testers make n tests, the chance that the BMD will alter at least one of the votes on one of the test ballots—and thus that parallel testing would detect the problem—is $1 - (1 - p)^n$.

For $p = 0.01$ (i.e., a 1% chance that the BMD alters votes on each ballot, independently across ballots), marking $n = 50$ test ballots would give a 39% chance of catching the BMD. To have a 95% chance of catching the problem, the sample size would need to be $n = 300$ test ballots.

I understand that a BMD can handle about 140 voting transactions on election day. Testing enough to have a 95% chance of detecting a 1% problem, i.e., 300 tests, would leave no time for voters to use the BMD. Even for pre-election L&A testing, where capacity for actual voters is not an issue, conducting 300 tests would take about 25 hours.

(Wallach 2019) argues that considering $p = 0.01$ is unrealistic, that a malicious actor would always attempt to “cheat” by a large amount (p rather greater than 1%). But an attacker seeking to avoid detection would not try to alter more votes than necessary (with some room for error). Changing votes on 1% of ballots can alter the margin of a jurisdiction-wide contest by 2%, if there are no undervotes, and by more than 2% if there are undervotes.⁵

Many contests are decided by less than that. Indeed, the margin in the 2016 U.S. presidential election was 0.22% in Michigan, 0.37% in Rhode Island, 0.72% in Pennsylvania (one of the states poised to buy BMDs to serve all voters), and 0.76% in Wisconsin. If the results had been different in Michigan, Pennsylvania, and Wisconsin Trump would have received fewer than 270 electoral votes, the number required to win.

Moreover, altering the votes on 1% of ballots overall can change the margin of smaller contests by far more than 2%. For instance, if a contest is only on 10% of the ballots,

⁴ Those who claim that L&A testing is a reliable means of discovering malware, bugs, and misconfiguration do not seem to have accounted for the vast number of combinations of things that could trigger a problem, as illustrated below.

⁵ The usual way of reporting margins ignores undervotes: it is the difference between the number of votes the winner and runner-up received, divided by the number of valid votes in the contest. The “diluted margin” is the difference between the number of votes the winner and runner-up received, divided by the total number of *ballots* cast in the contest, rather than the total number of *votes* cast in the contest. The diluted margin is never larger than the margin, but it can be much smaller. Changing the votes on $x\%$ of the ballots can change the diluted margin by up to $2x\%$, but can change the ordinary margin more. For instance, if the undervote rate is 30%, then changing votes on 1% of the ballots can change the margin by $0.02/0.7 = 2.9\%$.

altering votes on 1% of all ballots could change the margin in that particular contest by 20% if there are no undervotes. If the undervote rate in the contest is 30%, altering the votes on 1% of all ballots could change the margin in that particular contest by nearly 29%.

The sample size calculations above are based on an unrealistic model for how a malicious adversary would alter votes. An adversary has better strategies to avoid detection than these pro forma calculations—and those in (Wallach 2019)—assume. We now explore some of the options available to an attacker, illustrating that effective parallel testing requires sample sizes that are impossible in practice.

3 What is the threat model?

Suppose Alice, Bob, Carol, and Dan are running against each other in a plurality (whoever gets the most votes wins) contest. A malicious actor, Mallory, wants Alice to win. Mallory is not going to alter ballots at random. Mallory is not going to turn votes for Alice into undervotes or votes for a different candidate; it would only make Mallory’s job harder. Mallory would need to alter even more votes, which would increase the risk of detection. Rather, Mallory is going to turn some undervotes or votes for other candidates into votes for Alice.

Is Mallory going to cause every machine to alter some votes? Only some machines? Only machines in precincts where Alice is already expected to win, to keep the precinct totals unsuspicious? Will the programming cause the machine to alter votes all day, or only, say, in the last hour of voting, or when the lines for voting are predicted to be longest (when testers might be reluctant to conduct tests, because it would slow down voting)? Will the programming only alter votes if the voter is using the audio interface?⁶ Only if the voter takes a long time to vote?⁷ Only if the voter changes selections repeatedly?⁸

Will Mallory hack the BMD to make every vote for other candidates equally likely to be altered into a vote for Alice? Or will Mallory make the machine alter votes only when some other pattern of votes occurs in other contests on the ballot, e.g., only when the voter does not make any selections in some set of contests?⁹

To have a large chance¹⁰ of detecting a particular problem, testers need to use “enough” tests under the conditions that trigger the vote alterations. That might mean testing particular machines using particular vote pattern or patterns at particular times of day when the machines have been used in a particular way, and using the machines in a particular way.¹¹

The malware might or might not change votes at random: if we knew what the malware did, we wouldn’t need testing. To talk about the probability of detecting a problem, we

⁶ That might signal that the voter is less likely to check the printed selections.

⁷ That might signal that the voter is elderly or has trouble reading.

⁸ That might signal that the voter is less likely to remember the final selections, and more likely to assume that any discrepancy is their own fault.

⁹ That might signal that the voter is not very interested in those contests, and thus might not be alert to changes.

¹⁰ Here, “chance” is meant in a non-technical sense: see below.

¹¹ For instance, using the audio interface, changing the font size, displaying the ballot in a language other than English, taking a short or a long amount of time to vote, etc. The possible strategies are endless.

need to assume that the *testers* are testing at random. For instance, testers could test randomly selected machines at times randomly selected during the day, using randomly selected test patterns. They could choose randomly whether to use the touchscreen interface, the audio interface, or the sip-and-puff interface; whether to change the font size, whether to display the ballot in a language other than English; whether to vote slowly or quickly; whether to revise selections and how many times; and so on.

3.1 Attacks using vote patterns and time of day

There is an enormous number of possible voting patterns. For instance, if there are 5 contests, each with 4 candidates, there are $5^5 = 3125$ possible vote patterns, including the possibility of undervotes. In our hypothetical example where the contest between Alice, Bob, Carol, and Dan is one of the 5 contests on the ballot, there are $4 \times 5^4 = 2500$ patterns that do not already have a vote for Alice. Mallory will alter votes when the voter selects some subset of those patterns, possibly at random, possibly only at particular times of day, and possibly only when the BMD is being used in a particular way, e.g., when it is busiest or when the voter takes a long time. The malware might never alter votes if the BMD has been idle for more than 5 minutes, on the assumption that testing is more likely to take place when the polling place is “calm.” The malware might never alter votes unless the user takes more than 10 minutes to mark the ballot, on the assumption that testers will be in a hurry and will only rarely tie up a BMD for that long for a single test. There is a very large space of strategies available to Mallory.

To illustrate the difficulty of detecting problems, suppose that Mallory alters *every* ballot on which the voter selected Bob, but *only* if the voter also selects the candidate from Bob’s party in the next two contests on the ballot. And Mallory only alters votes for 30 minutes a day—at the time predicted to be busiest, say 12:15-12:45pm.

Suppose that 10% of voters vote during that half hour. Since many people vote along party lines, Mallory might expect a large fraction of voters who choose Bob will pick someone in Bob’s party in the next two contests, too. For simplicity, let’s suppose everyone who votes for Bob does just that.

Suppose that 10% of all voters vote during that half hour, and that 40% of those voters vote for Bob (and for the candidates in Bob’s party in the two following contests). Then Mallory will flip votes on $10\% \times 40\% = 4\%$ of the ballots, changing the margin between Alice and Bob by 8% for ballots marked using *that* BMD. Hacking 25% of all BMDs in the same way would change the overall margin by 2%.

3.2 What is the chance Mallory gets caught?

Suppose the jurisdiction has 400 BMDs in all, of which 100 were hacked.

In the scenario described above, tests of the 75% of machines that were not hacked will not find any error whatsoever. Any tests using 60% of vote patterns that don’t trigger the hack will not find any error whatsoever. Any tests run outside the critical 12:15–12:45 window will not find any error whatsoever.

Since that is (by assumption) the busiest time, the testers might not run any tests at all in that half hour. Let’s assume that they run one test between 12:15 and 12:45, on an affected machine, using a random test pattern.

What is the chance that any of the tests finds a discrepancy? Of the 3125 possible test patterns, only $5^2 = 25$ will trigger the malware. That is, there is only a $25/3125 = 0.8\%$

chance that the testers will catch that BMD, no matter how many times they test it outside the window 12:15–12:45.

Suppose every machine is tested in the same way, using test patterns selected independently at random, and all happened to be tested between 12:15 and 12:45. Even though 100 machines have been infected and the margin has been changed by 2%, the chance that *none of the tests of any of the 400 machines* finds a discrepancy is $(1 - 0.008)^{100} = 45\%$.

This strategy for altering votes is *not* a worst-case scenario, just one of countless possible attacks. In particular, this attack mentioned does not use information about BMD settings or voter interaction with the BMD.

3.3 Attacks using BMD settings and voter interaction

BMDs are a perfect device to disenfranchise voters with disabilities, older voters, and voters whose preferred language is not English. That is because the BMD “knows” whether the voter is using the audio interface or the sip-and-puff interface, whether the voter has changed the font size, whether the BMD is displaying the ballot in English, how long the voter is taking to vote, and whether the voter is changing selections or reviewing ballot pages. It can therefore be programmed—deliberately or accidentally—to alter the selections of just such voters.

Thus, parallel tests need to probe not only vote patterns and time of day, but also BMD settings and voter interactions with the BMD: anything known before the election or that the BMD can monitor on election day can be used to target votes to alter.

For instance, suppose that 5% of voters take more than 2 minutes per page to vote. Then a substantial number of the live tests in every polling place need to take more than 2 minutes per page. Otherwise, if the malware only alters votes when the voter takes longer than 2 minutes, it could change some margins by up to 10% with no possibility of detection.¹² Note that this means testing cannot (always) be fast. Gilbert’s proposed testing protocol¹³ assumes that tests will take about 4 minutes each, and does not contemplate the possibility that an attack could be targeted using the duration of the voting session. The attack described here has no possibility of being detected by his proposed protocol.

Similarly, suppose that 5% of voters change at least two of their selections. Then a substantial number of the live tests in every polling place need to change at least two selections. Otherwise, if the malware only alters votes if the voter changes two or more selections, it could change some margins by 10% with little or no possibility of detection.

Similarly, if 0.8% of voters use the audio interface,¹⁴ a substantial number of tests in every polling place need to use the audio interface. Otherwise, if malware only alters votes when the voter uses the audio interface, it could change some margins by 1.6% with little or no possibility of detection.

¹² In this example and the ones that follow, the maximum change depends in part on what the true voter preferences in each contest are. For instance, if all such voters voted for Bob, changing their votes to votes for Alice would alter the margin by 10%. If votes in that group were evenly split between Alice and Bob, changing the votes for Bob into votes for Alice would alter the margin by 5%. If none of them voted in the contest in question, changing their votes to votes for Alice would alter the margin by 5%. If all of them voted for Alice in the first place, no change to those votes could increase Alice’s margin.

¹³ <http://www.juangilbert.com/BallotMarkingVerificationProtocol.pdf>, last visited 18 August 2019.

¹⁴ Approximately 0.8% of the U.S. population is legally blind. <https://www.nfb.org/resources/blindness-statistics>, last visited 15 August 2019.

If 25% of voters display the ballot in a language other than English,¹⁵ a substantial number of tests in every polling place need to set the BMD to use a language other than English. Otherwise, the BMD could change some margins by 50% with little or no possibility of detection.

Live tests need to probe *every* subset of voter preferences, BMD settings, and voter interactions with the BMD that could alter *any* contest outcome, and they need to probe every such subset enough to have a high probability of detecting any changes to selections in that subset.

For illustration, imagine a single contest expected to be neck-and-neck across all four candidates, Alice, Bob, Carol, and Dan. Mallory knows from past experience that about 5% of voters in those precincts will display the ballot in English, take more than 10 minutes to vote, and change at least two of their selections before printing, and that this behavior is not associated with voter preferences.¹⁶

Mallory rigs the machines that contain the contest identically: they will all randomly flip votes for Bob into votes for Alice with 50% probability, but only if the voter displays the ballot in English, takes more than 10 minutes to vote, and changes at least two selections. That will make Alice's margin over Bob 2.5%, and her margin over Carol and Dan 1.25%.

To have a 95% chance of finding this problem, there need to be 5 tests in those precincts where the testers display the ballot in English, select Bob in that contest, take more than 10 minutes to vote, and change their selections at least twice ($1 - (1 - 0.5)^5 = 0.968$). This alone would tie up BMDs for more more than 50 minutes. Since testers would not know ahead of time that Mallory is trying to rig the election in favor of Alice, to protect just that single contest against the same hack in favor of the other candidates, they would also need to do those tests but with votes for Alice, votes for Carol, votes for Dan, and undervotes. That would tie BMDs up for $5 \times 50 = 250$ minutes, i.e., 4 hours and 10 minutes.

Of course, Mallory might have hacked the BMDs to change votes for Carol into votes for Alice, but only for voters who display the ballot in English, take 5-7 minutes to vote, and do not change any selections. If voters who display the ballot in English, take less than 5-7 minutes to vote, and do not change any selections comprise 10% of voters, then Mallory would only have to flip 25% of the votes for Carol in that group to votes for Alice to get the same margins. To have 90% chance of detecting the problem, there need to be 11 tests in those precincts where testers selected Carol, took 5-7 minutes to vote, and did not change any selections ($1 - (1 - 0.25)^{11} = 0.958$). Those 11 tests take at least 55 minutes, and again, since the testers would not know in advance that Mallory was using votes for Carol to benefit Alice, they also need to do those tests, using votes for Alice, for Bob, for Dan, and undervotes. That would tie BMDs up for $5 \times 55 = 275$ minutes, i.e., 4 hours and 35 minutes.

Because the conditions that trigger those two attacks are mutually exclusive, testing for one does not test for the other: we have to test for both. This brings us to 8 hours and

¹⁵ In 2013, roughly 26% of voters in Los Angeles County, CA, spoke a language other than English at home. http://rrcc.lacounty.gov/VOTER/PDFS/PUB/2013_Multilingual_Access_Elections.pdf, last visited 15 August 2019.

¹⁶ Such things could be measured by malware on BMDs deployed in previous elections, or possibly inferred from BMD internal logs from previous elections. It is not likely to be something that an election official knows or could measure legally or ethically: taking such measurements would likely compromise vote anonymity.

45 minutes of testing, just for these two possible attacks on that single contest, among many other possible attacks.

For instance, Mallory might instead have changed votes for Bob into votes for Alice, but only when the voter displays the ballot in a language other than English and takes 10 minutes or more to vote. If such voters comprise 10% of all voters, the calculation two paragraphs above shows that to have a 95% chance of detecting this attack requires at least another 4 hours and 35 minutes of testing. We are now at 13 hours and 20 minutes of testing.

But we have not exhausted the possible attacks, even on that one contest. For instance, Mallory could have hacked the BMDs to change votes for other candidates into votes for Alice only under *other*, mutually exclusive conditions, which could depend on *other* variables as well, such as votes in other contests, time of day, recency of the last voter's session, BMD load, etc. The 13 hours of tests already described cannot not reveal such hacks—even in principle—because the attacks involve mutually exclusive conditions.

And all of this testing is just for one contest. Once additional contests are on the ballot, there need to be tests for attacks against them, too. In practice, live testing cannot provide much protection against malware.

4 A little math

4.1 Mallory's goal

Let i denote a set of conditions that could apply to a voter's interaction with a BMD, including patterns of possible preferences in the elections on the ballot, time of day, duration of voting session, BMD user settings such as language and font size, whether the voter revised selections and which were revised, etc. Let f_i be the frequency with which actual voter sessions meet those conditions; we assume there is a complete enumeration of conditions, so that $\sum_i f_i = N$, the total number of ballots marked by voters. To alter the margin between Alice and Bob by m votes, Mallory needs to alter votes on a fraction $\pi_i \in [0, 1]$ of the sessions that satisfy the set of conditions i , in such a way that $\sum_{i \in \mathcal{I}} \pi_i f_i \geq m/2$. (The bound is sharp only if all the conditions i for which $\pi_i > 0$ already have a vote for Bob and not for Alice; otherwise, more than $m/2$ ballots need to be altered.)

Suppose that, given the fraction π_i , Mallory picks the particular $\pi_i f_i$ sessions that satisfy condition i to alter randomly, independently of anything the tester does, and independently across i . Then, if the tester tries condition i , the conditional probability that Mallory will alter any votes in that session—and hence that the tester will detect that Mallory interfered—is π_i .

Suppose the testers know $\{f_i\}$, i.e., they know how many sessions satisfy each set of conditions.¹⁷ If the testers test condition i with probability $f_i/(\sum_j f_j)$, the probability

¹⁷ If this were literally true, the testers would already know the correct outcome of every contest, and thus could detect any interference by Mallory simply by looking at the overall totals: there would be no need to audit—or even to conduct the election. Moreover, to estimate or measure f_i would require capability that is not currently (supposed to be) part of voting systems, because it would compromise vote anonymity. Mallory could, however, estimate f_i well using malware previously installed on BMDs.

that a given test will reveal Mallory's attack is

$$\frac{\sum_i \pi_i f_i}{\sum_i f_i} \geq \frac{m}{2N}.$$

In general Mallory will be able to alter outcomes by changing votes randomly for sessions that share some set of characteristics that any given auditing strategy is unlikely to probe enough to detect.

4.2 Auditing as an adversarial game

The auditing problem can be viewed as a two-player game, tester versus Mallory: the tester picks an auditing strategy, which must be public. With knowledge of that strategy, Mallory picks a (possibly random) rule for altering votes to in a way that has a large chance of changing the outcome of one or more contests. If Mallory can keep the probability that the audit will find any changed votes low, Mallory wins. If the auditing strategy ensures that if Mallory altered the outcome of one or more contests, the audit has a large chance of sounding an alarm, the tester wins.

Nothing in this paper should be construed to imply that parallel tests *only* need to detect malicious attacks. Bugs, misconfiguration, and other problems can also alter votes. Parallel testing does not solve the BMD security problem unless it has a sufficiently large chance of detecting outcome-changing faults, whatever their cause.

4.3 The curse of dimensionality

Several features of this problem make random testing require large sample sizes to be effective:

1. The testers do not know which contest(s) may be affected, nor which candidate(s) in those contests.
2. The outcome of a small contest or a contest with a narrow margin can be changed by altering only a small fraction of votes—depending on actual voter preferences, an arbitrarily small fraction of votes.
3. Malware or errors can be triggered by a vast number of combinations of many variables.

The high dimension requires exponentially more sampling, in general. As a statistical problem, auditing election outcomes using a trustworthy paper trail (e.g., checking outcomes using a risk-limiting audit (Stark 2008) (Stark 2009) (Lindeman and Stark 2012)) is easier than parallel testing of BMDs, in part because such audits only need to detect errors that favored the reported winner(s), and in part because whether anything affected the tabulation of the cast ballots is evident by looking at the cast ballots—provided the paper trail is trustworthy. In contrast, parallel BMD testing needs to detect errors that favor *any* candidate and to catch the malware in the act, by trying the right input to the right BMD at the right time.

4.4 An oracle bound

Suppose the tester could ask an oracle whether a particular cast BMD printout had an error. This would obviate the need even to know what variables characterize voters'

interactions with the BMD, much less to know how often voters interact with the BMD in any particular way (vote preferences, time of day, speed of voting, interface, language, etc.).

Suppose a contest was on the ballot on 20 BMDs, which printed a total of $20 \times 140 = 2800$ ballots on election day. Suppose that the BMDs altered 14 of the ballots, which could change the outcome of the contest in question by 1% if there were no undervotes, and by more if there were undervotes. (For instance, if the undervote rate is 50%, then changing votes on 0.5% of the ballots could change the margin by 2%.)

The tester will ask the oracle whether a set of randomly selected cast BMD printouts had any errors. How big would that set need to be in order to have at least a 95% chance of finding at least one printout with an error? The answer is the smallest value of n such that

$$\frac{2800 - 14}{2800} \cdot \frac{2799 - 14}{2700} \cdots \frac{2800 - (n - 1) - 14}{2800 - (n - 1)} \leq 0.05,$$

i.e., $n = 539$ voters, about 20% of the capacity of those machines. This would mean testing each BMD about 27 times on election day, on average: several times an hour, even under these counterfactual, optimistic assumptions.

We can turn the oracle bound around to ask how large a contest would need to be so that oracle-based testing would have at least a 95% chance of detecting a change to 0.5% of the ballots cast in the contest using not more than 13 tests per day per machine, on average. (That corresponds to testing a machine once per hour for a 13-hour day, and to using roughly 9.2% of BMD capacity for testing rather than voting.) The answer is that it would need to be on at least 47~BMDs, i.e., at least 6,580 votes, even under these impossibly optimistic assumptions.

But in reality, testers cannot look over voters' shoulders to see whether the BMD printed what it was supposed to. Testers do not know what variables characterize voters' interactions with BMDs, nor how often votes fall into any range of values of those variables: they could not measure these things without collecting data that would compromise the anonymity of votes. There would need to be more tests to compensate for that lack of information—many more. In reality, tests have to be *in addition to* actual voters' use of the BMDs, so the jurisdiction would need more machines (at least 24 machines instead of 20 in the first example; at least 52 instead of 47 in the second example) to accommodate the testing. In reality, contests can be decided by less than 1% of the ballots, so there would need to be more checks. And in reality, some contests are on fewer machines, which also increases the amount of testing required.

5 Complications

5.1 The only remedy is a new election

If testing catches a BMD altering votes, it is clearly appropriate to remove that BMD from service and conduct a forensic investigation of that BMD—and all the other BMDs used in the jurisdiction.¹⁸ But what about *this* election? There is no way to figure out which ballots were affected, nor even how many ballots were affected. There is not even a way to tell how many BMDs were affected, especially because malware could erase itself after election day, leaving no forensic evidence.

¹⁸ It's also appropriate to contact DHS and the FBI.

Thus, there is no way to know whether the election outcome reflected by the printed output is the same as the outcome according to what the BMDs told voters on confirmation screens (or through audio confirmation). Conducting a new election is the only way to rectify the damage.

5.2 Margins are not known

Margins are not known until the election is over, when it is too late to do more testing if contests have narrower margins than anticipated. Testing to a pre-defined threshold, e.g., to ensure a 95% chance of detecting errors in 0.5% or more of the votes in any contest, will sometimes turn out not to suffice. See below.

5.3 Tests have uncertainty

Relying on parallel testing to detect BMD faults has intrinsic uncertainties that should not be neglected.

For instance, suppose it were possible to design practical parallel tests that had a 95% chance of sounding an alarm if BMDs alter 0.5% or more of the votes in any contest. A reported margin of 1% or less in a plurality contest¹⁹ would be below the “limit of detection”²⁰; morally, the outcome of such a contest would be a tie. Would we revise election law to require automatic runoff elections whenever a reported margin is below 1%? If not, elections cease to be democratic and become an arbitrary means of deciding political contests.

The principle of evidence-based elections (Stark and Wagner 2012) says that election officials should not only report the winners, but also provide persuasive evidence that the reported winners really won—or admit that they cannot, in which case there should be a new election. In this hypothetical, that would mean holding a new election if the reported margin were below 1%.

And if a reported margin is greater than 1% (or the relevant limit of detection), we would still have only 95% “confidence” that BMD faults (bugs, misconfiguration, or hacking) did not cause the wrong candidate to appear to win. If the BMD output reflects the wrong electoral outcome, a perfect full manual tally, recount, or risk-limiting audit based on BMD printout will (with high probability) confirm that wrong outcome.

5.4 Asymmetry of information

The testing strategy will be public: it is a matter of regulation or law. But the attacker’s strategy is private. Thus, an attacker can adapt to the announced testing strategy, but the attacker does not have to announce its strategy. Even if the tests contain random or adaptive elements, testing must be limited to a small number of tests per machine per day, to leave time for actual voting. That limits the ability of testing to find problems that affect only a small fraction of votes.

¹⁹ For non-plurality social choice functions such as ranked choice or instant runoff, there is more than one definition of the margin, and the margin can be difficult to calculate. The relevant question is whether a change to 0.5% of the votes in a contest could cause the wrong candidate to appear to win, given the social choice function in that contest. If so, then the contest outcome is in doubt, and a new election seems like the appropriate remedy.

²⁰ If there are undervotes, the “limit of detection” measured as a margin would be even higher, as discussed above. That is, changing 0.5% of the votes could change the margin by more than 1%.

5.5 Malicious attacks are not the only risk

Misconfiguration, malfunction, and bugs can also alter votes. One cannot assume, as (Wallach 2019) does implicitly, that the only problems parallel testing needs to detect are those that a clever opponent would engineer.

5.6 Randomness is key

The test patterns need to be unpredictable, or malware could simply leave the votes alone if the input pattern matches a known or likely test pattern. Similarly, the test times need to be unpredictable (not, for instance, during the first 5 minutes of each hour), or malware could simply leave votes alone when tests are expected.

It is impractical to test at uniformly distributed random times, or after a random number of voters have used each BMD. For instance, testers will probably not mark 5 test ballots in a row on the same BMD. More likely, the sampling plan would involve testing hourly or at other regular intervals. For instance, Gilbert suggests testing once an hour (<http://www.juangilbert.com/BallotMarkingVerificationProtocol.pdf>), which makes it easier for malware to evade the test. If the polling place is busy when the test is supposed to occur, there will be pressure on the testers not to slow down voting by testing the BMD. If there is less testing when BMDs are busy, it's easier for malware to avoid the test.

5.7 New systems and extra hardware would be needed

Parallel testing requires new infrastructure to generate the random test patterns “just-in-time” and to signal the testers to conduct tests at random times. Infrastructure to log and report test results and to respond to discrepancies would also be needed.

Parallel testing also requires deploying *more* BMDs, because some (or all) of the capacity will be taken up by the live testing, reducing (or completely exhausting) the capacity for actual voters.

5.8 Extra staff and training would be needed

Who will conduct these tests? Pollworkers or elections staff would need additional training. Tests would probably need to involve two people, to ensure that the tester did not make a mistake in inputting selections. Additional pollworkers or other staff would be needed in every polling place to perform the additional work. The most important time to test is when the polling place is busiest, because that's when the most votes could be affected in the shortest time. Thus, effective testing is almost guaranteed to slow down voting.

6 Can the number of spoiled ballots be used to detect problems?

(Wallach 2019) also suggests comparing the rate of ballot “spoilage” to the expected background rate as a statistical test for BMD malfunction. Again, while this idea has heuristic appeal, numbers matter.

The Poisson distribution is a standard model for independent rare events. Without worrying about whether it is a good model for voting errors, let's use it as a guide to intuition.

Suppose that the background rate of ballot spoilage, based on previous elections, is 1%; that is, about 1% of voters request a fresh ballot. We suppose that the spoilage rate is the same for all machines in all precincts. We also suppose that about 100 voters use each BMD on election day.

Imagine a contest that appears on 2,000 BMDs, collectively marking 200,000 ballots. If things are functioning properly, we would expect about $200000 \times 0.01 = 2000$ spoiled ballots. What observed number of spoiled would give 95% confidence that something unusual is going on? The answer is 2074: 2073 spoiled ballots would not sound an alarm. If 10% of voters whose selections are mis-recorded notice the error and spoil their ballots,²¹ that would correspond to 730 BMD errors. That means that malware, bugs, or other misconfiguration could alter the margin in this contest by $2 \times 730/200000 = 0.73\%$. But it could alter the margin in a smaller contest by far more without creating a statistically suspicious number of spoiled ballots.

The numbers depend strongly on the number of BMDs that contain the contest. For instance, if a contest includes 50 BMDs (5,000 voters), then under the same assumptions we would expect 50 spoiled ballots if everything is functioning correctly. It would take 62 spoiled ballots to sound an alarm at 95% confidence, i.e., a 5% chance of a false alarm if everything is actually working properly. Finding 61 spoiled ballots would not sound an alarm. The $61 - 50 = 11$ extra spoiled ballots signal 110 errors, which could shift the margin by $2 \times 110/5000 = 4.4\%$. Smaller contests can be altered by even larger amounts without raising an alarm.

Even under ideal circumstances, passive monitoring based on spoiled ballots does not produce direct evidence of malfunction; it does not identify which ballots, if any, have errors; and it does not provide any evidence about whether the errors changed any outcomes. It seems implausible that an election official would hold a new election simply because the number of spoiled ballots was a bit larger than expected.

7 Can BMDs ever be secure?

Current BMDs as currently deployed have an insurmountable security flaw described in (Stark 2019) and (Appel, DeMillo, and Stark 2019): the protocols make voters responsible for checking whether BMDs are performing correctly, voters cannot get any evidence to prove to others that a malfunction occurred. As a result, election officials will always be in a bind if voters complain of problems: there is no way to tell the difference between a misbehaving machine, voter error, and a cry of “wolf.”

Some protocols developed for end-to-end cryptographically verifiable (E2E-V) elections can allow voters to prove that machines misbehaved in particular ways.²² While those protocols do not apply to BMDs, similar ideas might make it possible for a voter to obtain evidence of MBD malfunction that could be presented to others. None has been proposed.

The design and use of BMDs needs to be re-thought from the ground up so that either

- a. voters get evidence of malfunctions that can be used to demonstrate to others that a malfunction occurred, or

²¹ The work of (DeMillo, Kadel, and Marks 2018) suggests this is very optimistic.

²² E.g., the “Benaloh challenge,” as described in (Benaloh 2006). The Benaloh challenge can provide proof that a machine encrypted a plaintext vote incorrectly, but it does not address what happens if the voting machine prints the wrong plaintext vote.

14 Philip B. Stark

- b. some other mechanism is in place that has a high probability of catching all BMD malfunctions that could alter election outcomes.

Parallel testing is an attempt at (b). Unfortunately, it requires far more testing than is possible in practice. “Passive” testing by monitoring the number of spoiled ballots is also an attempt at (b). Unfortunately, the signal-to-noise ratio makes passive testing ineffective, especially for smaller contests. In any event, the only remedy if any BMD is discovered (or suspected) to have altered votes is a new election. Voting systems should instead be designed to be *resilient*, so that they can recover from errors without re-running the election. This is inherent in “strong software independence” (Rivest and Wack 2006).

8 Conclusion

While in theory parallel testing or logic-and-accuracy testing might help protect against malfunctioning or hacked BMDs, in practice, there are so many attack strategies (and possibilities for subtle bugs) that no reasonable amount of testing could guarantee even a modest chance of finding an outcome-changing problem. An attack (or bug) could depend on any number of variables, including, among others:

- the current voter’s selections (including undervotes) in every contest
- selections made by previous voters who used that BMD
- time of day
- BMD load
- BMD user settings, such as font size, ballot language, whether the audio interface is in use, whether the sip-and-puff interface is in use
- voter interaction with the BMD, such as how quickly the voter completes each page, whether and how often the voter revises selections, and whether the voter chooses to ignore/override undervote warnings

Because there are so many possible outcome-changing problems that involve mutually exclusive sets of conditions and different contests, it’s effectively impossible to protect against malware or subtle bugs by parallel testing or logic and accuracy testing. The number of tests required to get a reasonable level of assurance is simply too high.

Moreover, even if it were possible to get, say, 95% confidence that no more than 0.5% of the votes were changed in any contest, fairness would then demand a runoff election in any (plurality) contest with a diluted margin²³ of 1% or less.²⁴

No jurisdiction has conducted parallel testing of this kind. Any parallel testing requires more BMDs (to leave enough capacity for actual voters), new infrastructure, new procedures, and more pollworker training. Testing when the polls are busiest is especially important—and especially disruptive. The sample sizes required for testing to be effective against outcome-changing errors are prohibitively large (the same is true for logic and accuracy testing): considerable time and labor are required. And if a problem is detected, the only safe response is to hold a new election: there’s no way to tell which votes were changed, nor even how many votes were changed.

²³ See footnote 5.

²⁴ For non-plurality social choice functions such as ranked choice or instant runoff, the margin can be quite difficult to calculate. The relevant question would be whether a change to 0.5% of the ballots could cause the wrong candidate to appear to win, whatever the social choice function. If so, then the contest outcome is in doubt, and a new election seems like the appropriate remedy.

Absent some clever new design for a BMD-like device that enables voters to provide public evidence of any malfunctions they experience—without making elections vulnerable to false accusations and without compromising voter privacy—there is no way to ensure that BMDs have not misprinted enough votes to change election outcomes.

Nonetheless, well-designed BMDs²⁵ are arguably the best technology for voters with some kinds of disabilities to mark and cast a paper ballot independently. To my knowledge, existing commercial BMDs do not provide a means for blind voters to check whether the printed output matches their intended selections. It would not be technically challenging to add “verification stations” that speak a printed ballot to the voter, so the voter can spoil the ballot and start over if there is a discrepancy.

If BMDs are deployed, they should undergo some pre-election logic and accuracy testing to screen for gross configuration errors. BMD printout should be designed to make it as easy as possible for voters to check whether the printout matches their intended choices. Voters who use BMDs should be urged to check whether the printout matches their intended choices, and to request another chance to mark a ballot if there’s a discrepancy. Voters with disabilities should be provided an accessible way to check BMD printout. And election officials should pay attention to the number and distribution of spoiled BMD printouts.

But for the foreseeable future, prudent election administration also requires keeping the number of ballots marked by machines as small as possible—while ensuring that every voter is provided a means to mark, verify, and cast a paper ballot independently.

Bibliography

- Appel, A.W., R.A. DeMillo, and P.B. Stark. 2019. “Ballot-Marking Devices (Bmds) Cannot Assure the Will of the Voters.” https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3375755.
- Benaloh, J. 2006. “Simple Verifiable Elections.” *Proceedings of the USENIX/Accurate Electronic Voting Technology Workshop and Electronic Voting Technology Workshop* Vancouver, Canada.
- DeMillo, R., R. Kadel, and M. Marks. 2018. “What Voters Are Asked to Verify Affects Ballot Verification: A Quantitative Analysis of Voters’ Memories of Their Ballots.” <https://ssrn.com/abstract=3292208>. <https://doi.org/http://dx.doi.org/10.2139/ssrn.3292208>.
- Lindeman, M., and P.B. Stark. 2012. “A Gentle Introduction to Risk-Limiting Audits.” *IEEE Security and Privacy* 10: 42–49.
- Perez, E. 2019. “Georgia State Election Technology Acquisition: A Reality Check.” https://trustthevote.org/wp-content/uploads/2019/03/06Mar19-OSETBriefing_GeorgiaSystemsCostAnalysis.pdf; OSET Institute Briefing.

²⁵ There are many poorly designed BMDs on the market, including designs with blatant, easily exploited security flaws. See, e.g., <https://tyt.com/stories/4vZLCHuQrYE4uKagy0oyMA/5YIEQxHW5qmWayG0kYCsy2> or <https://freedom-to-tinker.com/2018/09/14/serious-design-flaw-in-ess-expressvote-touchscreen-permission-to-cheat/> (both visited 20 August 2018). Moreover, recent testing in Pennsylvania suggests that some current BMDs do not enable voters with disabilities to vote independently (Secretary of the Commonwealth of Pennsylvania 2018, pp68–90).

16 Philip B. Stark

Rivest, R.L., and J.P. Wack. 2006. “On the Notion of Software Independence in Voting Systems.” <http://vote.nist.gov/SI-in-voting.pdf>.

Secretary of the Commonwealth of Pennsylvania. 2018. “Report Concerning the Examination Results of Election Systems and Software EVS 6012 with DS200 Precinct Scanner, DS450 and DS850 Central Scanners, ExpressVote HW 2.1 Marker and Tabulator, ExpressVote XL Tabulator and Electionware EMS.”

Stark, P.B. 2008. “Conservative Statistical Post-Election Audits.” *Annals of Applied Statistics* 2: 550–81. <http://arxiv.org/abs/0807.4005>.

———. 2009. “Risk-Limiting Post-Election Audits: P -Values from Common Probability Inequalities.” *IEEE Transactions on Information Forensics and Security* 4: 1005–14.

———. 2019. “Ballot-Marking Devices (BMDs) Are Not Secure Election Technology: Letter to Subcommittee on Voting Technology of Government Affairs Committee, Georgia House of Representatives.” <https://www.stat.berkeley.edu/~stark/Preprints/bmd19.pdf>.

Stark, Philip B., and David A. Wagner. 2012. “Evidence-Based Elections.” *IEEE Security and Privacy* 10: 33–41.

Wallach, D. 2019. “On the Security of Ballot Marking Devices.” <https://arxiv.org/pdf/1908.01897.pdf>.



Political Science

School of Public & International Affairs

UNIVERSITY OF GEORGIA

GEORGIA VOTER VERIFICATION STUDY

JANUARY 22, 2021

[DRAFT VERSION 1.0]

AUDREY A. HAYNES

(polaah@uga.edu)

M.V. HOOD III

(th@uga.edu)

I. Study Objective

In 2019, the State of Georgia made the decision to replace its aged direct-recording electrical (DRE) voting system with a new system produced by Dominion Voting Systems that would be utilized across the state.¹ For in-person voting (early or on election-day) this new voting system provided ballot marking that was electronic but also provided a paper ballot output. Voters make selections on a touch-screen and may review their selections before printing a paper ballot from a printer attached to the ballot marking device. The printed ballots contain a QR code that can be scanned and also displays in printed words all the selections made by the voter. Before leaving the voting booth, voters are encouraged to review the choices on their printed ballot and, if there are discrepancies, they may spoil their ballot and vote again. When ready, voters physically take their printed ballot and insert it into a ballot tabulator. At this point the ballot is officially recorded as being cast and the paper ballot is secured within the tabulator.

The statewide rollout for the new system was to have taken place during the 2020 presidential preference primary. Due to the Covid-19 pandemic, this primary was canceled and rescheduled to coincide with the May statewide primary election. Almost half of primary voters in May voted absentee by mail and, consequently, did not have any interaction with the new system. For these reasons, our study was delayed until the 2020 general election.

The purpose of the study is to examine the interaction of Georgia voters with the new ballot marking system. More specifically, we are seeking to determine if voters check the choices registered on their paper ballot before depositing (casting) them in the tabulator.

II. Study Design

In order to study the interaction of Georgia voters with the new Dominion ballot marking voting system we recruited University of Georgia students to act as field observers at various precinct locations. Student observers were compensated and were required to attend a one-hour in-person training session. Seventy-four observers were assigned a precinct location and worked a six-hour shift, either 7:00 am to 1:00 pm or 1:00 pm to 7:00 pm on election-day (November 5, 2020).²

Students were trained to record various observations related to voters and voter-interaction at their assigned precinct. All observers used a common observation sheet [a copy of the observation sheet is attached to this report]. Students were instructed to collect specific observations for each voter including the time they entered the voting booth and the time they completed the process. Basic demographic information was also collected using the following categories:

- Age [Young, Middle, Older]³
- Sex [Male, Female, Unsure]
- Race [White, Black, Hispanic, Asian, Unsure]

At the voting booth, observers were to record whether the voter checked/reviewed the choices on their printed ballot and, if so, the approximate amount of time they spent in this process.

¹Dominion ImageCast x Prime Ballot Marking Device (Version 5.5.10.30).

²These students were given credential letters from the Secretary of State's Office and worked with poll managers to optimize their observations while also maintaining all rules governing public safety protocols and voter privacy.

³Observers were instructed to use the following age ranges as approximations for these categories: Young (18-29); Middle (30-54); and Older (55+).

Observers were trained to mentally note the number of seconds the voter examined their paper ballot using the following categories:

- Brief Glance (Less than one second)
- Short Time (One to five seconds)
- Long Time (More than five seconds)

Students were also instructed to observe voter behavior at the tabulator where a precinct official was stationed. Per instructions issued from the Secretary of State's Office, these polling place officials were to oversee the process of voters inserting their ballots into the tabulator. They were also asked to remind voters to check the choices on their ballot before casting it, if they wished. At this stage of the process student observers recorded whether or not the precinct worker instructed the voter to check their ballot and, second, if the voter was provided instructions on how to cast their ballot using the tabulator. We also recorded whether or not the voter checked over their ballot while waiting to insert it into the tabulator.

Thirty-one precincts were selected at random from the following counties: Clarke (17), Oconee (5), Oglethorpe (2), Madison (2), Jackson (1), and Barrow (5).⁴ These six counties are comprised of the City of Athens (Clarke) and the immediate geographically proximate counties. In order to assure racial/ethnic diversity among these counties, a stratified random sample of thirty precincts was randomly selected from the 44 precincts that are majority non-Hispanic white in composition.⁵ A further three precincts were randomly selected from the group of five majority-minority precincts. In addition, a random sample of seven precincts was also drawn from Gwinnett County. Gwinnett County is part of the Atlanta metro area. Of these, six were majority-minority and one was majority non-Hispanic white [A list of precincts where observations were taken is appended to this report].

This study also included an embedded experiment where voters at four randomly selected precincts were given a paper reminder (see below) to check their ballot after printing.⁶ The reminder was patterned directly on a public relations campaign from the Secretary of State's Office. Upon completion of the observational study, we can compare voters at treatment (received reminder) and control precincts (did not receive reminder) in order to determine if handing individuals a reminder will produce any perceivable effect on voter behavior.

⁴In order to ensure a steady stream of observations precincts with fewer than 2,000 registrants were not considered for random selection.

⁵Race/ethnicity of registrants is recorded in Georgia. Precinct racial composition was obtained from the Georgia Secretary of State.

⁶The four precincts where voters received a reminder are as follows: Covenant Life Sanctuary (4)-Barrow County; Judia J. Harris Elementary School (2A)-Clarke County; Athens Transit Multi-Modal Center (4A)-Clarke County; and Bishop Baptist Church (4&5)-Oconee County.



After printing your paper ballot,
REVIEW your choices
to make sure they are correct.

III. Analysis and Findings

Following the general election, observation sheets were collected and data were entered digitally in preparation for analysis. In total, there were 4,024 usable observations. The following discussion of findings relies on the aforementioned observation data.

Voter Behavior at the Voting Booth

This section discusses the behavior of voters at the voting booth after printing their paper ballot. The results of our analysis indicate that 80.1% of the voters observed checked their ballot, compared to a fifth (19.9%) who did not.

Given that some voters in the *brief glance* category only undertook a very cursory examination of their ballot, a fairer assessment on the question of whether voters checked the choices on their paper ballot is found in Table 1. Here we see that just under a third of the voters observed (31.3%) only glanced at their ballot after printing it.

Combining this group with those who failed to look at their ballot at all, we see that more than half (51.3%) of voters undertook an insufficient check of their ballot from the standpoint of being able to adequately determine if their choices had been correctly documented on the paper ballot. On the other hand, approximately half (48.7%) the voters observed did spend at least some time examining their ballot. Approximately three of ten voters (29.9%) spent a short time looking over the choices on their paper ballot, while close to one-fifth (18.8%) spent a longer period of time examining their ballot.

Table 1. Voter Behavior Observed at the Voting Booth

Category	Percentage
Did not check	20.0%
Brief glance	31.3%
Short time	29.9%
Long time	18.8%

Voter Behavior at the Tabulator

In this section we examine voter behavior as voters moved from the voting booth to deposit their ballot in the tabulator.⁷ At this stage of the process, student observers recorded that 23.8% of voters were instructed to check their ballot by precinct workers. This would appear far short of the goal set by the Secretary of S that all voters be informed at this stage in the process that they should examine the choices on their paper ballot before casting it. On the other hand, almost all voters (86.7%) were given instructions by precinct workers on inserting their ballot into the tabulator.

Of the cases where observers were able to hear voters being instructed to check their ballots, 62.3% did so at this point in the voting process, while 37.7% did not. This 62%, however, equates to only 14.2% of the total number of voters observed in the study. Most voters then did not check their ballot at this stage in the voting process.⁸ Of those voters who did check over their ballots before the tabulator, 85.7% had also been recorded as having done so at the voting booth. (85.7%). Only 80 voters, or 2.0% of the total number of voters observed, checked their ballots at the tabulator stage having not done so at the voting booth.

Voter Behavior and Demographics

In this section we will discuss the demographic breakdown of the voters in our observation study, and how our primary observation of interest (checking one's ballot at the voting booth) varied by demographic category. The breakdown of race, sex, and age for the observation set is detailed in Table 2 below.

⁷Student observers were instructed to position themselves at the precinct where they could observe voter behavior both at the voting booth as well as the tabulator. In some locations, this was difficult to achieve and, as a consequence, it was not possible to always track voter behavior at the tabulator. The results presented in this section are based on the data that are available.

⁸It should be noted that in most precincts it was a very short distance from voting booths to the tabulator. Those voters who did check their ballot in the voting booth, therefore, may have opted out of checking their ballot again only seconds later.

Table 2. Demographic Breakdown of Voters Observed

	Percentage
Race:	
White	67.4%
Black	21.9%
Hispanic	5.9%
Asian	2.9%
Unknown	1.9%
Sex:	
Male	49.4%
Female	50.2%
Unknown	0.4%
Age:	
Young	37.1%
Middle	42.9%
Older	18.4%
Unknown	1.6%

Next, we provide a series of contingency tables in order to determine if the decision of voters to check their ballot at the voting booth varied by race, sex, and age.⁹ For the contingency tables presented, we collapse the *Did not check* and *Brief glance* categories into a single category and the *Short time* and *Long time* categories into a single category. Looking at Table 3, white voters were slightly more likely to have checked their ballot as compared to minority voters. The difference between white and black voters on this metric is 6.0 points (50.5% of white voters were recorded as having checked their ballot versus 44.5% of black voters). The difference between white and Hispanic voters is 4.9 points and for Asian voters it is 6.2 points. The distribution of voters who checked their ballot across racial categories in Table 3 is statistically significant, an indication the denoted pattern is not due to random chance.

Table 3. Voter Behavior by Race

Category	White	Black	Hispanic	Asian
Did not check/glance	49.5%	55.5%	54.4%	55.7%
Short/long time	50.5%	44.5%	45.6%	44.3%

Notes: Entries are column percentages. Chi-square: 11.35 (p<.01)

Table 4 displays the same information by Sex. Here we see a bare majority (50.5%) of male voters checked over their ballots at the voting booth compared to 46.6% of female voters. The difference between these two groups, at 3.9 points, is statistically significant.

⁹Voters who could not be classified by observers (unknown) are excluded from these analyses.

Table 4. Voter Behavior by Sex

Category	Male	Female
Did not check/glance	49.5%	53.4%
Short/long time	50.5%	46.6%

Notes: Entries are column percentages. Chi-square: 5.90 ($p < .05$)

Finally, Table 5 examines the results by age category. Here we see that the likelihood of checking one's ballot at the voting booth is positively associated with age. Less than a majority of younger voters (45.3%) were observed checking their ballots, compared with 49.2% of middle-aged voters, and 53.2% of older voters. The difference between younger and older voters is 7.9 points. The distribution of responses presented in Table 5 across age categories is statistically significant.

Table 5. Voter Behavior by Age

Category	Young	Middle	Older
Did not check/glance	54.7%	50.8%	46.8%
Short/long time	45.3%	49.2%	53.2%

Notes: Entries are column percentages. Chi-square: 12.67 ($p < .01$)

To summarize, the profile of a voter who checked their ballot at the voting booth was associated by demographic characteristics. White voters, male voters, and older voters were all more likely to have spent time checking their paper ballots after printing.

Embedded Experiment

Table 6 below compares the behavior of voters in the treatment and control precincts based on whether they checked their ballot at the voter booth after printing. There were a total of 511 voter observations recorded at the four treatment precincts, or 13.1% of all observations. Those voters in precincts that received a reminder (treatment) were 5.3 points more likely to have checked their ballot as compared to voters that did not receive a reminder (control), 53.2% versus 47.9% respectively. Although five points is a modest increase, the difference between the treatment and control groups is statistically significant.¹⁰ Translated into English, the simple act of giving voters a paper reminder at the check-in table increases the likelihood that they will check over their ballot after printing.

Table 6. Results of Embedded Experiment

Category	Treatment	Control
Did not check/glance	46.8%	52.1%
Short/long time	53.2%	47.9%

Notes: Entries are column percentages.

¹⁰Mean difference: -0.528; $t = -2.228$; $p < .05$.

IV. Conclusion

On election-day for the 2020 general, a largescale observational study was undertaken in north Georgia to study the interaction of voters with the state's new ballot marking system. More specifically, we were tasked with observing whether or not voters checked to confirm the choices on their paper ballot after printing. Using an objective measure, we found that half of the voters observed spent some or a long period of time checking over their paper ballot. If voters did check their paper ballot, it was at the voting booth, immediately after printing out their ballot. Few voters rechecked their ballots at the tabulator and almost no voters checked their ballot for the first time at the tabulator stage in the process. This finding may, in part, be due to the fact that poll workers stationed at the tabulator did not consistently appear to be reminding voters to do so, focused instead on the process of inserting the ballot correctly and explaining the process to the voter. These same poll workers did, however, adequately instruct almost all voters on how to cast their ballot at the tabulator. Providing voters a reminder of their ability to confirm their choices on the paper ballot at the tabulator stage should be reemphasized during precinct poll worker training. Distinctive demographic characteristics based on age, sex, and race were noted for voters who checked their ballot at the voting booth. Finally, the results of our embedded experiment demonstrate that handing out a simple reminder to voters can act to modify voter behavior—in this case modestly increasing the probability they will check their ballot choices after printing.

Appendix A: Observation Sheet

GEORGIA VOTING BEHAVIOR CODE SHEET

Coder name: _____

Date: _____ Start Time: _____ County: _____ Precinct: _____ Assigned Voter Booths: _____

Start Time/ Stop Time	Sex	Race	Age	Voter Behavior at Voting Booth	Length of check [in seconds]	Voter Behavior at Tabulator [Check all that apply]	Other Observations: [Required poll worker assistance at voting booth; requested new ballot, etc.]
	Male Female Unsure	White Black Hispanic Asian Unsure	Young [18-29] Middle [30-54] Older [55+] Unsure	Checked ballot Did not check Unsure	Brief glance <1 Short time 1-5 Long time >5 Unsure	__Instructed to check ballot __Checked ballot __Did not check ballot __Instructed to insert ballot Other obs. [note in next column]	
	Male Female Unsure	White Black Hispanic Asian Unsure	Young [18-29] Middle [30-54] Older [55+] Unsure	Checked ballot Did not check Unsure	Brief glance <1 Short time 1-5 Long time >5 Unsure	__Instructed to check ballot __Checked ballot __Did not check ballot __Instructed to insert ballot Other obs. [note in next column]	
	Male Female Unsure	White Black Hispanic Asian Unsure	Young [18-29] Middle [30-54] Older [55+] Unsure	Checked ballot Did not check Unsure	Brief glance <1 Short time 1-5 Long time >5 Unsure	__Instructed to check ballot __Checked ballot __Did not check ballot __Instructed to insert ballot Other obs. [note in next column]	
	Male Female Unsure	White Black Hispanic Asian Unsure	Young [18-29] Middle [30-54] Older [55+] Unsure	Checked ballot Did not check Unsure	Brief glance <1 Short time 1-5 Long time >5 Unsure	__Instructed to check ballot __Checked ballot __Did not check ballot __Instructed to insert ballot Other obs. [note in next column]	
	Male Female Unsure	White Black Hispanic Asian Unsure	Young [18-29] Middle [30-54] Older [55+] Unsure	Checked ballot Did not check Unsure	Brief glance <1 Short time 1-5 Long time >5 Unsure	__Instructed to check ballot __Checked ballot __Did not check ballot __Instructed to insert ballot Other obs. [note in next column]	

Appendix B: Precincts Included in Study

Precincts used in Study

County	Precinct ID	Polling Place	Address	Observer	Shift
BARROW	1	BETHLEHEM COMMUNITY CENTER	750 Manger Avenue Bethlehem, GA 30620	1	7 am to 1 pm
				2	1 pm to 7 pm
BARROW	4	COVENANT LIFE SANCTUARY	115 Patrick Mill Rd SW Winder, GA 30680	1	7 am to 1 pm
				2	1 pm to 7 pm
BARROW	5	FIRE STATION 1 (STATHAM)	1625 Bethlehem Road Statham, GA 30666	1	7 am to 1 pm
				2	1 pm to 7 pm
				3	1 pm to 7 pm
BARROW	13	WINDER COMMUNITY CENTER	113 E. Athens Street Winder, GA 30680	1	7 am to 1 pm
				2	7 am to 1 pm
				3	1 pm to 7 pm
BARROW	16	THE CHURCH AT WINDER	546 Treadwell Road Bethlehem, GA 30620	1	7 am to 1 pm
				2	1 pm to 7 pm
CLARKE	1A	WINTERVILLE TRAIN DEPOT	125 NORTH CHURCH ST., WINTERVILLE GA 30683	1	1 pm to 7 pm
				2	7 am to 1 pm
CLARKE	1D	WHIT DAVIS ELEMENTARY SCHOOL	1450 WHIT DAVIS RD., ATHENS GA 30605	1	1 pm to 7 pm
				2	7 am to 1 pm
CLARKE	2A	JUDIA J. HARRIS ELEM. SCHOOL	2300 DANIELSVILLE RD., ATHENS GA 30601	1	1 pm to 7 pm
				2	7 am to 1 pm
				3	1 pm to 7 pm
				4	7 am to 1 pm
CLARKE	3B	THOMAS N. LAY PARK	297 HOYT ST., ATHENS GA 30601	1	7 am to 1 pm
CLARKE	4A	ATHENS TRANSIT MULTI-MODAL CTR	775 E BROAD ST., ATHENS GA 30601	1	7 am to 1 pm
				2	1 pm to 7 pm
CLARKE	4B	MEMORIAL PARK	293 GRAN ELLEN DR., ATHENS GA 30606	1	7 am to 1 pm
				2	12 to 6 pm
CLARKE	5A	OGLETHORPE AVENUE ELEMENTARY SCHOOL	1150 OGLETHORPE AVE., ATHENS GA 30606	1	7 am to 1 pm
				2	1 pm to 7 pm
CLARKE	5B	WHITEHEAD RD ELEM SCHOOL	555 QUALWOOD DR., ATHENS GA 30606	1	7 am to 1 1:30
				2	7 am to 1 pm
				3	1 pm to 7 pm
				4	1 pm to 7 pm
CLARKE	5D	ACC FLEET MGMNT BLDG	225 NEWTON BRIDGE RD., ATHENS GA 30607	1	7 am to 1 pm
				2	7 am to 1 pm
				3	1 pm to 7 pm
CLARKE	6A	CLEVELAND RD ELEM SCH	1700 CLEVELAND RD., BOGART GA 30622	1	7 am to 1 pm
				2	1 pm to 7 pm
CLARKE	6C	TIMOTHY RD ELEM SCHL	1900 TIMOTHY RD., ATHENS GA 30606	1	1 pm to 7 pm
CLARKE	7A	UNITARIAN UNIVERSALIST FELLOWSHIP	780 TIMOTHY RD., ATHENS GA 30606	1	1 pm to 7 pm
CLARKE	7B	ATHENS REGIONAL LIBRARY	2025 BAXTER ST., ATHENS GA 30606	1	7 am to 1 pm
				2	1 pm to 7 pm

Precincts used in Study

County	Precinct ID	Polling Place	Address	Observer	Shift
CLARKE	7C	FIRESTATION #3	1198 SOUTH MILLEDGE AVE., ATHENS GA 30605	1	7 am to 1 pm
				2	1 pm to 7 pm
CLARKE	8A	GAINES ELEMENTARY SCHOOL	280 GAINES SCHOOL RD., ATHENS GA 30605	1	1 pm to 7 pm
CLARKE	8B	CEDAR SHOALS HIGH SCHOOL	1300 CEDAR SHOALS DR., ATHENS GA 30605	1	7 am to 1 pm
				2	1 pm to 7 pm
CLARKE	8C	FIRESTATION #7	2350 BARNETT SHOALS RD., ATHENS GA 30605	1	7 am to 1 pm
				2	1 pm to 7 pm
JACKSON	3 (Central Jackson)	HOPE CROSSINGS CHURCH	2106 Old Pendergrass Rd, Jefferson, GA 30549	1	7 am to 1 pm
				2	7 am to 1 pm
				3	1 pm to 7 pm
MADISON	Danielsville	MADISON COUNTY BOARD OF ELECTIONS	94 Spring Lake Drive Danielsville, GA 30633	1	1 pm to 7 pm
				2	7 am to 1 pm
MADISON	Hull	HULL COMMUNITY CENTER	1346 Old Elberton Rd Hull, GA 30646	1	7 am to 1 pm
				2	1 pm to 7 pm
OCONEE	1	WATKINSVILLE CITY HALL	191 VFW Drive Watkinsville, GA 30677	1	1 pm to 7 pm
				2	7 am to 1 pm
OCONEE	9	BOGART COMMUNITY CENTER	141 East Thompson St. Bogart, GA 30622	1	1 pm to 7 pm
				2	7 am to 1 pm
OCONEE	12	OCONEE COUNTY CIVIC CENTER	2661 Hog Mountain Rd. Watkinsville, GA 30677	1	1 pm to 7 pm
				2	1 pm to 7 pm
				3	7 am to 1 pm2
OCONEE	4 & 5	BISHOP BAPTIST CHURCH	1110 Old Bishop Rd. Bishop, GA 30621	1	1 pm to 7 pm
				2	7 am to 1 pm
OCONEE	6 &7	GRACE FELLOWSHIP CHURCH	1120 Malcom Bridge Rd. Bogart, GA 30622	1	7 am to 1 pm
				2	1 pm to 7 pm
OGLETHORPE	Beaverdam	SONLIGHT BAPTIST CHURCH	1134 Hargrove Lake Road Colbert, Georgia 30628	1	7 am to 1 pm
				2	1 pm to 7 pm
OGLETHORPE	Crawford	CRAWFORD DEPOT	1158 Athens Road Crawford, Georgia 30630	1	7 am to 1 pm
GWINNETT	88	BETHANY BAPTIST CHURCH	2306 Bethany Church Road Snellville GA 30039	1	7 am to 1 pm
GWINNETT	59	LANDMARK CHURCH	3737 Holcomb Bridge Road Norcross GA 30092	1	1 pm to 7 pm
GWINNETT	57	B B HARRIS ELEM SCHOOL	3123 Claiborne Drive Duluth GA 30096	1	1 pm to 7 pm
GWINNETT	123	IGL BAUTISTA NUEVA JERUSALEN	5300 Williams Road Norcross GA 30093	1	1 pm to 7 pm
GWINNETT	155	EPIPHANY LUTHERAN CHURCH	1350 Peachtree Industrial Blvd Suwanee GA 30024	1	1 pm to 7 pm
GWINNETT	51	IGLESIA METROPOLITANA SDA CHURCH	5990 Oakbrook Parkway Norcross GA 30093	1	7 am to 1 pm
GWINNETT	119	KNIGHT ELEMENTARY	401 North River Dr Lilburn GA 30047	1	7 am to 1 pm

September 8, 2022

Georgia State Elections Board
Mr. William S. Duffey, Jr., Chair
Mr. Matthew Mashburn, Member
Mrs. Sara Tindall Ghazal, Member
Mr. Edward Lindsey, Member
Dr. Janice W. Johnston, Member
Secretary of State Brad Raffensperger, Ex Officio
214 State Capitol
Atlanta, Georgia 30334

Dear Chair Duffey and Members of the State Elections Board:

Media reports have recently confirmed allegations that Georgia's voting system software was accessed and copied by several unauthorized individuals. These individuals handled the sensitive files in a reckless manner, transferring them to numerous people over the internet, who also have no authority to possess the state's voting software and data.¹

As members of the computer science, cybersecurity, and election integrity communities,² we are writing to provide important context regarding the serious threats this security breach poses to Georgia's elections, and to urge you to address the issue by taking specific actions to mitigate the heightened risks.

The immediate concern

The Secretary of State claims that his office has adequately addressed this security breach by replacing one server in the one currently known affected county. Replacing the server does not mitigate the breach, for reasons we shall explain.

The illegal copying of software and data ("disk images") from the Georgia election management system (EMS) and voting device software, which occurred more than

¹ Emma Brown, Jon Swaine, Aaron C. Davis, Amy Gardner, "Trump-allied lawyers pursued voting machine data in multiple states, records reveal," *The Washington Post*, August 15, 2022. Available at: <https://www.washingtonpost.com/investigations/2022/08/15/sidney-powell-coffee-county-sullivan-strickler/>

² The undersigned are all experts in election cybersecurity. Each of us has well over a decade of continuous experience in that field and a long history of conducting technical studies of voting systems or voting system-related cybersecurity, as well as writing, speaking, testifying, making media appearances on many aspects of election integrity.

20 months ago, constitutes a serious threat to Georgia’s election security. Those images, which include the EMS, and its installation environment, were accessed improperly by individuals and entities that have engaged in a campaign aimed at overturning the 2020 election results in Georgia and other key states.³ While it is prudent to assume that other nation states have had that software for a long time, now countless individuals with unknown affiliations, motives, and physical access to voting systems have it also. This increases both the risk of undetected cyber-attacks on Georgia, and the risk of accusations of fraud and election manipulation. Without trustworthy, physical records of voter intent (recorded by the hand of the voter, not with vulnerable, computerized ballot marking devices), rigorous chain of custody of ballots, and rigorous post-election auditing, such allegations will be difficult, if not impossible, to disprove.

Georgia’s elections have a higher risk of compromise due to the reliance on computerized Ballot-Marking Devices (BMDs) to record vote selections for in-person voting.

To ensure resiliency, auditability and transparency in an election, it is essential that there be a reliable, trustworthy record of each voter’s selections; this provides ground truth of voter intent.⁴ This record should be the record used for audits and recounts. Having a trustworthy physical record of voter intent allows administrators to check and confirm that the vote tabulation is correct, and to catch and correct any errors that may have occurred—regardless of their source.

In 2020 Georgia finally abandoned its insecure, paperless, touchscreen Diebold voting machines. Ignoring recommendations from election security experts⁵ and public preference,⁶ the Secretary of State successfully pushed the State legislature to approve a universal use BMD touchscreen voting system. The BMD system has put Georgia’s elections at a higher risk for tampering, disruption, or allegations of

³ Tierney Sneed, “Judge sanctions pro-Trump lawyers who brought ‘frivolous’ lawsuits,” *CNN*, August 26, 2022. Available at: <https://www.cnn.com/2021/08/25/politics/judge-sanctions-powell-wood-kraken-lawsuits/index.html>

⁴ “Report of the Auditability Working Group,” National Institute of Standards and Technology, 2010. Available at: <https://www.nist.gov/document/auditabilityreportxml-7htm>

⁵ Dr. Wenke Lee, the only computer security expert on Secretary Raffesperger’s “Secure, Accessible, Fair, Elections (SAFE) Commission, vigorously opposed the universal use of ballot marking devices. Dr. Lee’s position was supported by 24 computer security experts who urged the SAFE Commission to recommend against the universal use of BMDs.

⁶ Mark Neisse, “AJC poll: Georgians support paper ballots and oppose voter purges,” *Atlanta Journal Constitution*, January 21, 2019. Available at: <https://www.ajc.com/news/state--regional-govt--politics/ajc-poll-georgians-support-paper-ballots-and-oppose-voter-purges/mkdeIgUXtzJL6TFVbM6BVP/>

manipulation because all votes cast in a polling location are recorded using vulnerable, computerized BMDs, and tabulated from QR codes, which voters cannot check. The software breach in Coffee County has further increased this risk.

Vulnerabilities in the Ballot Marking Device (BMD) software can be exploited to mis-record votes.

While serving as an expert witness in the *Curling v. Raffensperger* lawsuit in federal court in Georgia, University of Michigan computer science professor J. Alex Halderman, one of the nation's foremost experts in voting system cybersecurity, analyzed Dominion ICX BMD (touchscreen and printer) software. Dr. Halderman found serious security vulnerabilities, some of which would allow a voter to infect a BMD with malware while voting, with little likelihood of detection. That malware could make the BMD print incorrect votes and spread silently to other voting machines and the central election management system in the county. Halderman's findings confirm that Dominion ICX BMD printout is not a reliable record of voter intent.⁷

The judge in *Curling* considered Prof. Halderman's full report, dated July 1, 2021, so sensitive that she ordered the report to be sealed. Halderman's follow-on report, dated July 13, 2021, is public and summarizes some of the conclusions of this sealed report.⁸ Prof. Halderman's findings were so concerning that he presented them to the Department of Homeland Security's Cybersecurity and Infrastructure

⁷ Research shows that voters rarely check machine-printed votes and rarely notice errors when they do check. No audit can determine whether ballot-marking devices printed voters' true selections: if a substantial number of voters use ballot-marking devices, no audit can limit the risk that an incorrect electoral result will be certified. See, e.g., Appel, A., R.A. DeMillo, and P.B. Stark, 2020. Ballot-Marking Devices Cannot Ensure the Will of the Voters, *Election Law Journal: Rules, Politics, and Policy*, 19, <https://doi.org/10.1089/elj.2019.0619>; Seventh Declaration of Philip B. Stark, 13 September 2020. *Curling et al. v Raffensperger et al.*, United States District Court for the District of Georgia, Northern Division 1:17-cv-2989-AT <https://coalitionforgoodgovernance.sharefile.com/share/view/s5ae19303763c45dfa5c8238cb58e47d8> (last visited 2 September 2021); Eighth Declaration of Philip B. Stark, 2 August 2021. *Curling et al. v Raffensperger et al.*, United States District Court for the District of Georgia, Northern Division 1:17-cv-2989-AT <https://coalitionforgoodgovernance.sharefile.com/share/view/sbda3c49bc6b646579d6691fb68f2d840> (last visited 2 September 2021)

⁸ Declaration of J. Alex Halderman, 2 August 2021. *Curling et al. v Raffensperger et al.*, United States District Court for the District of Georgia, Northern Division 1:17-cv-2989-AT <https://coalitionforgoodgovernance.sharefile.com/d-s7d96b021c2d3419984512b56ff6eee95> (last visited 2 September 2021)

Security Agency (CISA) which issued an advisory warning of the security vulnerabilities.⁹

Still, the Secretary of State has inaccurately assessed the security threats to the BMDs by repeatedly claiming that an attacker could only corrupt one device at a time.^{10, 11} This is incorrect. CISA's vulnerability assessment confirmed Dr. Halderman's findings that there are several vulnerabilities that could be exploited to spread malware from device to device, increasing the impact of an attack.¹² The Secretary's failure to appreciate the gravity and urgency of this breach further imperils Georgia's elections.

Emergency measures can be taken to secure the election and maintain voter confidence.

This newly heightened risk can be mitigated by critical but straightforward action.

First, Georgia should immediately discontinue the universal use of the Dominion ICX BMD for in-person voters, and instead provide voters with emergency hand-marked paper ballots to be tabulated by the current system's optical scanners. Georgia state election rules currently require:

*The Superintendent shall cause every polling place and advance voting location to have a sufficient number of blank paper ballots that can be marked by pen available for use in the event of emergency. The election superintendent shall also be prepared to resupply polling places with emergency paper ballots in needed ballot styles in a timely manner while voting is occurring so that polling places do not run out of emergency paper ballots.*¹³

State rules explicitly direct election officials to prepare for the use of emergency paper ballots, marked by pen. This means all election administrators and pollworkers should *already* be trained in the distribution and use of paper ballots.

⁹ ICS Advisory (ICSA-22-154-01) Vulnerabilities Affecting Dominion Voting Systems ImageCast X, June 3, 2022. Available at: <https://www.cisa.gov/uscert/ics/advisories/icsa-22-154-01>

¹⁰ Mark Neisse, "Handling of Georgia election breach investigation questioned," Atlanta Journal Constitution, September 4, 2022. Available at: [Questions surround handling of election breach investigation in South Georgia county \(ajc.com\)](https://www.ajc.com/story/news/politics/elections/2022/09/04/georgia-election-breach-investigation-questioned/8252870002/)

¹¹ Emma Hurt, "What's going on with Coffee County?", *Axios*, September 7, 2022. Available at: [2020 election investigation puts a spotlight on Coffee County, GA - Axios Atlanta](https://www.axios.com/2020-election-investigation-puts-a-spotlight-on-coffee-county-ga-68d1e1e1-2022-09-07)

¹² See *supra* note 9.

¹³ Georgia Rule 183-1-12-.01 Conduct of Elections, Available at: <https://rules.sos.ga.gov/gac/183-1-12>

Georgia counties can scale up their existing procedures to shift the bulk of in-person voting to paper ballots marked by pen. These ballots can be counted by the tabulators currently in use: no new equipment, programming or training is needed.

Voters who need or prefer to mark a ballot with assistive technology can continue to use BMDs with assistive technology. BMD units could also be used as a backup balloting unit if the polling place runs short of a specific ballot style printed ballot. Minimizing the use of the BMDs reduces the threat that BMD tampering can alter election results.

Second, we urge you to use your authority to mandate a *statewide post-election risk-limiting audit (RLA)* of the outcome for all contests on the ballot. Current SEB practice is that only one state-wide contest be audited, every other year; this is insufficient. This proposed audit should be done completely transparently, with citizen observation, under the auspices of local county election officials. Post-election auditing of the outcome requires a trustworthy paper trail of hand-marked paper ballots with limited use of machine-marked ballots.

If a cyberattack, misconfiguration, bug, or procedural lapse changes the outcome, a properly conducted RLA *based on trustworthy paper ballots* will correct the outcome (with high probability). If the election outcome is correct in the first place, the RLA will provide strong public evidence that it is, creating a “firewall” against litigation and disinformation seeking to discredit the outcome.

We believe it is important that a public commitment to rigorous post-election verification be made before Election Day. Otherwise, it may appear to be a partisan decision, and there may be calls for other kinds of “audits” that are neither scientifically grounded nor probative, but could spuriously undermine public confidence in the election. We urge you to take the lead on the auditing issue early and reassure Georgia voters that a thorough transparent audit will promptly follow the election and be completed prior to certifying the results.

In bringing our concerns about the recent Dominion software compromise to your attention we are not accusing Dominion of wrongdoing. Nor do we have evidence that anyone currently plans to hack Georgia’s elections. However, it is critical to recognize that the release of the Dominion software into the wild has measurably increased the risk to the real and perceived security of the election to the point that emergency action is warranted.

We are all willing to discuss any of these points with you or your staff, either in writing or by phone or videoconference. We would be happy to help swiftly design a straightforward, practical, transparent statewide RLA process that will be a model for how elections should be secured. We would like to be helpful in any way that you find useful to defend against the threats posed by the escaped Dominion code and newly discovered Dominion BMD vulnerabilities. Please do not hesitate to call on us to assist.

Yours truly,

Mustaque Ahamad, PhD.
Professor
School of Cybersecurity and Privacy
Georgia Institute of Technology*

Duncan Buell, PhD.
Chair Emeritus — NCR Chair in Computer Science and Engineering
Department of Computer Science and Engineering
University of South Carolina*

Richard DeMillo, PhD.
School of Cybersecurity and Privacy
Charlotte B. and Roger C. Warren Chair in Computing
Georgia Institute of Technology*

Larry Diamond, PhD.
Senior Fellow, Hoover Institution and Freeman Spogli Institute,
Stanford University*

Lowell Finley
Former California Deputy Secretary of State for Voting System Technology and
Policy (2007-2014)

Susan Greenhalgh
Senior Advisor for Election Security
Free Speech For People

David Jefferson, PhD.
Lawrence Livermore National Laboratory* (retired)
Board of Directors, Election Integrity Foundation*

Douglas W. Jones, PhD.
Emeritus Associate Professor of Computer Science
University of Iowa*

Daniel P. Lopresti, PhD.
Professor, Department of Computer Science and Engineering*
President, International Association for Pattern Recognition (IAPR)*
Vice Chair, Computing Research Association's Computing Community
Consortium (CCC)*
Lehigh University*

Mark Ritchie
Former Secretary of State of Minnesota*
Past President National Association of Secretaries of State*

John E. Savage, PhD.
An Wang Professor Emeritus of Computer Science
Brown University*

Kevin Skoglund
President and Chief Technologist
Citizens for Better Elections*

Philip B. Stark, PhD.
Professor, Department of Statistics
University of California, Berkeley*

**Affiliations below are provided for identification purposes only. The statements and opinions expressed here are not necessarily those of our employers or institutions.*

September 2, 2021

Dr. Shirley N. Weber, Secretary of State
Office of the Secretary of State
1500 11th St.
Sacramento, California 95814

URGENT: Critical New Risks to the Recall Election Can Be Mitigated by the California Secretary of State

Dear Secretary Weber:

As you know, about three weeks ago, binary images of the Dominion election management system (EMS) were made public. While the software versions are not identical to those used in California, differences are relatively minor: the release materially elevates threats to the trustworthiness of the ongoing California recall election and to public trust in the election. We urge you to address the issue by taking one critical action – a statewide risk-limiting audit (RLA) of trustworthy paper ballots – which can substantially mitigate these threats.

The undersigned are all experts in election cybersecurity. Each of us has well over a decade of continuous experience in that field and a long history of conducting technical studies of voting systems or voting-related cybersecurity, as well as writing, speaking, testifying, making media appearances on many aspects of election integrity. Several of us have served on special panels and task forces appointed by previous California Secretaries of State and have worked closely with local election officials in California.

California has a long history of innovation and national leadership in election security. It was one of the first states to audit elections, with an audit law dating back to the 1960s. It was one of the first states to ban paperless voting systems as a result of a task force appointed by the Secretary of State. The 2007 California Top-To-Bottom Review (“TTBR”) was the first state-sponsored review of the security of voting systems and was a huge contribution toward understanding the security issues in all computerized voting systems. Risk-limiting audits (RLAs) were first developed as an outgrowth of a special working group appointed by the Secretary of State, also in 2007, and the first pilot RLAs were conducted in California counties with support from the Secretary of State. California was the second state to have legislation authorizing RLAs and is in the second year of a second set of pilot audits. And California was the first state to expose the dangers inherent in Internet voting, in the report of another Secretary of State’s task force in 2000, and was among the first states to ban Internet connections of any kind to its voting equipment. We are thus confident that California election officials are well positioned to take this next step of election security leadership at this critical time.

We are also fully aware of the numerous technical and procedural safeguards California employs to prevent many kinds of administrative errors and to defend against cyber-attacks of various kinds. Many of these were pioneered in California and we acknowledge and applaud them. The security concerns we are writing about here, however, are different from threats we have seen before, and cannot be fully defended against by technical means. The only strong defense against them is the statewide RLA we are recommending for the current election.

The immediate concern

The illegal public release about three weeks ago of binary images of the Dominion election management system (EMS) software and its installation environment constitutes a serious threat to the recall election. Two of the images came from Mesa County, Colorado, and one came from Antrim County, Michigan. Those images, which include the EMS and its installation environment, have been widely downloaded. While it is prudent to assume that other nation states have had that software for a long time, thousands of other people with unknown affiliations, motives, and physical access to voting systems now have it also. That increases the risk of undetected outcome-changing cyber-attacks on California counties that use Dominion equipment and the risk of accusations of fraud and election manipulation which, without rigorous post-election auditing, would be impossible to disprove. While the versions of the Dominion software that were released are not identical to the versions used in California, they are closely related, so this security breach imperils California elections.

Heightened risk of cyber-attack directed against Dominion counties in the recall election

Every complex software system has bugs and security flaws. Cybersecurity research has shown that election software has more than its share. Since that software is usually kept proprietary and secret, however, relatively few people have had the opportunity to examine, instrument, and test it closely enough to find exploitable flaws.

This is now no longer the case, at least with Dominion software. As of August 2021, thousands of unknown people can study the code and find weaknesses to plan attacks on elections. The attacks can be deployed by non-technical accomplices, including voters, building maintenance personnel, and election workers. Unfortunately, even extensive pre-election testing of the voting equipment may not deter or detect such attacks.

The Dominion software from Antrim County has been studied in detail recently by University of Michigan computer science professor Alex Halderman, one of the nation's foremost experts in voting system cybersecurity.¹ While serving as an expert witness in the Curling v. Raffensperger lawsuit in federal court in Georgia, Prof. Halderman found very serious security vulnerabilities in the Dominion Ballot Marking Device (BMD) system, some of which would allow an ordinary voter to insert malware into a BMD during a voting session, with little likelihood of detection. That malware could spread undetected to other voting machines and potentially to the central election management system (EMS) in the county.² That EMS software is now in the hands of countless unauthorized people after the Mesa and Antrim releases. Prof. Halderman's full report, dated July 1, 2021, is so sensitive that the Court in Curling v. Raffensperger ordered that it be sealed. We urge you to file a motion with Judge Totenberg to obtain a

¹ Dr. Halderman recently was engaged by Michigan Secretary of State, Jocelyn Benson, to conduct a review of the Antrim County, Michigan Dominion Voting System after human error caused vote count discrepancies in the November 2020 election. (https://www.michigan.gov/documents/sos/Antrim_720623_7.pdf)

² Declaration of J. Alex Halderman, 2 August 2021. Curling et al. v Raffensperger et al., United States District Court for the District of Georgia, Northern Division 1:17-cv-2989-AT <https://coalitionforgoodgovernance.sharefile.com/d-s7d96b021c2d3419984512b56ff6eee95> (last visited 2 September 2021). In his public declaration Dr. Halderman writes "Attackers could exploit these flaws [in Dominion code] to install malicious software, either with temporary physical access (such as that of voters in the polling place) or remotely from election management systems. I explain in detail how such malware, once installed, could alter voters' votes while subverting all the procedural protections practiced by the State, including acceptance testing, hash validation, logic and accuracy testing, external firmware validation, and risk-limiting audits (RLAs). Finally, I describe working proof-of-concept malware that I am prepared to demonstrate in court."

confidential copy of Prof. Halderman's sealed report to inform your cybersecurity team of the vulnerabilities he discovered.

In raising our concerns about the Dominion software release we are not accusing Dominion of wrongdoing. Nor do we have evidence that anyone currently plans to hack the recall election. However, it is critical to recognize that the release of the Dominion software into the wild has increased the risk to the security of California elections to the point that emergency action is warranted.

Emergency measure to secure the election and maintain voter confidence

This newly heightened risk can be mitigated by critical but straightforward action. We urge you to use your authority to mandate a *statewide post-election risk-limiting audit* of the outcome for the two questions on the recall ballot. RLAs have become the widely acknowledged gold standard of post-election auditing. This proposed audit should be done completely transparently, with citizen observation, and under guidance from your office (not vendors or third parties) and under the auspices of local county election officials to maintain Californians' strong voter confidence. RLAs of the outcome require a trustworthy paper trail of hand marked paper ballots with limited use of machine-marked ballots.³ At least 17 of California's 58 counties—of vastly different sizes and using a broad spectrum of voting systems from different vendors—have already conducted pilot RLAs, so the process is well understood by local election officials. Because the same two contests are on every ballot in the state, a RLA of the recall election is especially straightforward and efficient.

If an actual cyberattack silently changes the outcome of the election, or any other procedural or software error does, a properly conducted RLA based on trustworthy paper ballots will detect it and correct it (with high probability). If the election outcome is correct in the first place the RLA will provide strong public evidence that it is, creating a "firewall" against litigation and disinformation seeking to discredit the outcome.

We believe it is important that a public commitment to such post-election verification be made before Election Day. Otherwise, it may appear to be a partisan decision, and there may be calls for other kinds of "audits" that are neither scientifically grounded nor probative, and that would likely undermine public confidence in the election. We urge you as California's chief election official to take the lead on the auditing issue early and reassure California voters that a thorough transparent audit will promptly follow the election and be completed prior to certifying the results.

We are all willing to discuss any of these points with you or your staff, either in writing or by phone or videoconference or in person in Sacramento. We would be happy to help swiftly design a straightforward, practical, transparent statewide RLA process that will be a model for how high-profile elections should be secured. We would like to be helpful in any way that you find useful to defend against the threats posed

³ Research shows that voters rarely check machine-printed votes and rarely notice errors when they do check. No audit can determine whether ballot-marking devices printed voters' true selections: if a substantial number of voters use ballot-marking devices, no audit can limit the risk that an incorrect electoral result will be certified. See, e.g., Appel, A., R.A. DeMillo, and P.B. Stark, 2020. Ballot-Marking Devices Cannot Ensure the Will of the Voters, *Election Law Journal: Rules, Politics, and Policy*, 19, <https://doi.org/10.1089/elj.2019.0619>; Seventh Declaration of Philip B. Stark, 13 September 2020. Curling et al. v Raffensperger et al., United States District Court for the District of Georgia, Northern Division 1:17-cv-2989-AT <https://coalitionforgoodgovernance.sharefile.com/share/view/s5ae19303763c45dfa5c8238cb58e47d8> (last visited 2 September 2021); Eighth Declaration of Philip B. Stark, 2 August 2021. Curling et al. v Raffensperger et al., United States District Court for the District of Georgia, Northern Division 1:17-cv-2989-AT <https://coalitionforgoodgovernance.sharefile.com/share/view/sbda3c49bc6b646579d6691fb68f2d840> (last visited 2 September 2021)

by the escaped Dominion code and newly discovered Dominion BMD vulnerabilities. Please do not hesitate to call on us to assist. Our contact information is listed with our names below.

Sincerely, and best wishes,

All affiliations below are provided for identification purposes only. The statements and opinions expressed here are not necessarily those of our employers or institutions.

Mustaque Ahamad

Professor, School of Cybersecurity and Privacy, Georgia Tech
mustaq.ahamad@gmail.com

Duncan Buell

NCR Chair in Computer Science and Engineering (Emeritus), University of South Carolina
duncan.buell@gmail.com

Richard A. DeMillo

Charlotte B. and Roger C. Warren Professor of Computer Science and Chair, School of Cybersecurity and Privacy, Georgia Tech
rad@demillo.com

Candice Hoke

Founding Co-Director, Center for Cybersecurity & Privacy Protection, Cleveland-Marshall College of Law, Cleveland State University
shoke@icloud.com

Harri Hursti

Co-founder, Nordic Innovation Labs; Co-founder, Voting Village at DEFCON
harri@hursti.net

David Jefferson

Retired Computer Scientist, Lawrence Livermore National Laboratory
drjefferson@gmail.com

Wenke Lee

John P. Imlay Professor of Computer Science; Director Georgia Tech Cybersecurity Center; School of Cybersecurity and Privacy; Member, Georgia Commission on Safe Secure Elections
wenke.lee@gmail.com

Prof. Philip B. Stark

Professor, Department of Statistics, University of California, Berkeley
pbstark@berkeley.edu

December 12, 2022

The Honorable Merrick Garland
Special Counsel Jack Smith
U.S. Department of Justice
950 Pennsylvania Ave., N.W.
Washington, D.C. 20530-0001

The Honorable Christopher Wray
Assistant Director Robert Wells
Federal Bureau of Investigation
935 Pennsylvania Avenue
Washington, D.C. 20530-0001

The Honorable Jen Easterly
Director
Cybersecurity & Infrastructure Security Agency (CISA)
Department of Homeland Security
Washington, DC 20023

Dear Attorney General Garland, Special Counsel Smith, Director Wray, Assistant Director Wells, and Director Easterly,

Elections form the foundation for the legitimacy of government in the United States. Protecting the security of our elections constitutes a top national security priority, and we appreciate the work the Department of Justice and the Department of Homeland Security have done to address threats to election security.

We¹ are writing to you to call attention to significant multi-state events impacting ongoing election security which were first revealed in discovery in a civil lawsuit by a non-governmental plaintiff.² Because these events were revealed

¹ Free Speech For People is a non-profit, non-partisan public interest legal organization that works to renew our democracy and our United States Constitution for the people. As part of our mission, we are committed to promoting, through legal actions, secure, transparent, trustworthy and accessible voting systems for all voters. We are not parties or counsel in the civil litigation referenced in this letter. The coalition of signatories includes renowned computer security experts and election experts.

² Emma Brown, Jon Swaine, "Inside the secretive efforts by Trump allies to access voting machines," *The Washington Post*, October 28, 2022. Available at: <https://www.washingtonpost.com/investigations/2022/10/28/coffee-county-georgia-voting-trump/>

in a private lawsuit rather than through a law enforcement investigation, the significance and consequences may not have registered with the relevant federal agencies. Specifically, we are writing regarding the multi-state plan, directed and funded by attorneys for Donald Trump—including Sidney Powell, Lin Wood, and Jesse Binnall—to access voting systems and obtain and distribute copies of voting system software unlawfully,³ which could potentially constitute federal crimes and be relevant to investigations into efforts to overturn the 2020 presidential election.

The same software is used in voting systems across the U.S. Its misappropriation and distribution pose serious risks to the credibility and trustworthiness of election results. The matter urgently requires your agencies' attention. This alarming development was not addressed in the Public Service Announcement (PSA), published by CISA on October 4, 2022.⁴ On October 28, 2022, news reports confirmed that the Department of Homeland Security, the Federal Bureau of Investigation, the U.S. Capitol Police, and the National Counterterrorism Center released a joint intelligence assessment warning that violent domestic extremists pose a heightened threat to the security of the 2022 mid-term elections.⁵ Federal security agencies are rightly concerned about potential attacks on elections. But because the unauthorized access and copying of election system software was uncovered by private action, it appears that the consequences of those activities—orchestrated by individuals associated with domestic extremists—have not been factored into these risk assessments. We write to highlight this important information for your agencies to integrate it into threat assessments, criminal investigations, and risk mitigations.

1. Executive Summary

According to evidence obtained in a civil lawsuit, attorneys and allied advisors for Donald Trump directed and funded a plot that included copying voting system software from multiple jurisdictions in multiple states. The operatives successfully obtained software from both Dominion Voting Systems and Election Systems &

³ Jon Swain, Aaron C. Davis, Amy Gardner, Emma Brown, “Files copied from voting systems were shared with Trump supporters, election deniers,” *The Washington Post*, August 22, 2022. Available at: <https://www.washingtonpost.com/investigations/2022/08/22/election-system-copied-files-trump/>

⁴ Available at: <https://www.ic3.gov/Media/PDF/Y2022/PSA221004.pdf>

⁵ Geneva Sands, Sean Lyngass, “Feds warn that domestic violent extremists pose heightened threat to midterm elections,” *CNN*, October 28, 2022. Available at: <https://www.cnn.com/2022/10/28/politics/midterm-domestic-extremist-threat/index.html>

Software (ES&S).⁶ Together, ES&S and Dominion equipment and software count more than 70% of the votes in the U.S.⁷

Now that the Dominion and ES&S software has been accessed and distributed by individuals without authorization, the election system software is considered to be “in the wild,” available to an unknown number of people and organizations that could use the software to undermine, disrupt, or tamper with elections in a number of ways.⁸

Access to the software enables bad actors to install the software on their own hardware to create their own exact replicas of voting systems, probe them, and develop exploits. They can decompile the software to get the source code, study it for vulnerabilities, and develop malware tailored to the system. Such malware can be loaded onto systems by a poll worker, maintenance worker, or even a voter: little physical access is needed.⁹ Nor is Internet access needed. Bad actors could use the software to develop ways to cause the system to malfunction to prevent voters from voting or to manipulate the vote count.

Moreover, access to this software may also be used in service of disinformation campaigns to cast doubt on legitimate election results. Perhaps the most easily executed attack would be to simply use the software as the basis of fabricated evidence of election manipulation to delegitimize the results and destabilize our government.¹⁰

⁶ Emma Brown, Jon Swaine, Aaron C. Davis, Amy Gardner, “Trump-allied lawyers pursued voting machine data in multiple states, records reveal,” *The Washington Post*, August 15, 2022. Available at: <https://www.washingtonpost.com/investigations/2022/08/15/sidney-powell-coffee-county-sullivan-strickler/>

⁷ <https://verifiedvoting.org/verifier/#mode/navigate/map/makeEquip/mapType/normal/year/2022>

⁸ The threats to election security posed by the unauthorized distribution of voting system software are described in detail in the July 14, 2022 statement from the OSET Institute “Increasing Concerns About Amplified Threats To Voting Systems,” available at: https://trustthevote.org/wp-content/uploads/2022/07/14July22_StatementOnPublicDisclosures.pdf.

⁹ University of Michigan Professor J. Alex Halderman, one the nation’s foremost experts on election system security, examined the Dominion Voting System ballot marking device for the Curling lawsuit and documented multiple security vulnerabilities that could be exploited. Though his detailed report remains sealed by the Court, Prof. Halderman submitted a public declaration recounting a high level summary of his findings, which includes this statement, “*Attackers could exploit these flaws to install malicious software, either with temporary physical access (such as that of voters in the polling place) or remotely from election management systems.*” No. 17-cv-02989-AT (N.D. Ga. filed Aug. 8, 2017). Document 1304-3

¹⁰ We have already seen this play out. Matt DePerno, a former candidate to be Michigan’s Attorney General and prominent figure in denying the results of the 2020 election, used access to voting system software to support allegations of election fraud. See: “Michigan Candidate Matt DePerno Boasts of Effort Showing How to Stuff Ballots,” *Deadline Detroit*, September 3, 2022. Available at:

Insider attacks by temporary election workers are an increasing threat to election security. A statement issued by the Bipartisan Policy Center warned: “There is mounting concern that temporary election workers recruited and trained by organizations with nefarious intent may undermine security and trust in the election process.”¹¹ Since organizations with likely nefarious intent also now possess copies of voting system software, this creates a profound threat scenario.

Though some of the impacted states are reportedly pursuing individual criminal investigations,¹² the coordinated, multi-state plan by Powell indicates there could be potential federal criminal liability that compels intervention by the Department of Justice.

Furthermore, we ask that the Department of Justice’s National Security Division and the Department of Homeland Security’s Cybersecurity and Infrastructure Security Agency (CISA) urgently assess the increased threats to election security posed by the unauthorized distribution of voting system software to individuals who have already spread misinformation and may attempt to disrupt elections, and factor these threats into activities to protect elections.

Because the multistate plan was not discovered by local or state law enforcement (which presumably would have shared such findings with the Department of Justice through normal channels) but by the private civil litigation *Curling v. Raffensperger*, we are writing (1) to urge the Department of Justice and the Department’s Special Counsel to investigate this coordinated plot to copy election system software from multiple states; and (2) to urge the Department of Justice’s Counterterrorism Division and the Department of Homeland Security’s CISA to assess the ongoing threats introduced by the distribution of voting system software, and calibrate efforts to protect elections to include those risks.

https://www.deadlinedetroit.com/articles/31208/michigan_candidate_matt_deperno_boasts_of_effort_showing_how_to_stuff_ballots

¹¹ Grace Gordon, Rachel Orey, “Closing Gaps in Poll Worker Policy,” The Bipartisan Policy Center, October 3, 2022. Available at: <https://bipartisanpolicy.org/explainer/poll-worker-policy/>

¹² Nathan Layne, “Republican clerk could be charged in Michigan voting-system breach,” *Reuters*, October 4, 2022. Available at: <https://www.reuters.com/legal/exclusive-michigan-police-ask-prosecutors-consider-charging-republican-clerk-2022-10-04/>

2. Background

In 2017, a public interest organization and individual voters filed a lawsuit in federal court in Georgia challenging the use of the state’s voting equipment, *Curling v. Raffensperger*.¹³ The case, now in its second phase, is being heard by the Honorable Amy Totenberg in the Northern District of Georgia. (*Curling v. Raffensperger* does not allege any election has been incorrectly decided.) In the course of discovery, plaintiffs obtained evidence that certain individuals had accessed voting equipment in Coffee County, Georgia, and copied all election data and the software that records, counts, and reports votes.¹⁴ This software is used in all of Georgia’s 159 counties, making this a breach of the entire state’s software and system. (Similar versions of the software are used in other jurisdictions, including Riverside, California, and Washington County, Pennsylvania.) Plaintiffs sought additional discovery to ascertain what had happened in Coffee County.¹⁵

Through evidence obtained in discovery, the plaintiffs have established that the data management firm SullivanStrickler was engaged by Sidney Powell to go to the Coffee County, Georgia, elections office and image all Dominion voting system equipment and devices.¹⁶ Cathy Latham, an influential Georgia Republican Party official, and an “alternate elector,” represented by Sidney Powell and Lin Wood in then-pending litigation to overturn the 2020 election, was a key local Coffee County organizer of the software collection effort. The plaintiffs in *Curling* also obtained surveillance camera video that shows several individuals, including Doug Logan of the Cyber Ninjas, visiting the Coffee County election office over the course of several days for extensive equipment access.¹⁷ Plaintiffs also obtained copies of SullivanStrickler’s voting system equipment forensic images and system

¹³ No. 17-cv-02989-AT (N.D. Ga. filed Aug. 8, 2017). Affidavits and depositions referred to herein are available at this docket.

¹⁴ Jose Pagliery, “Texts Reveal GOP Mission to Breach Voting Machine in Georgia,” *The Daily Beast*, June 5, 2022. Available at: <https://www.thedailybeast.com/how-a-coffee-county-gop-chair-coordinated-a-voting-machine-breach>

¹⁵ Jose Pagliery, “Subpoenas Probe GOP Mission to Breach Georgia Voting System,” *The Daily Beast*, June 15, 2022. Available at: <https://www.thedailybeast.com/subpoenas-probe-gop-mission-to-breach-georgia-voting-system>

¹⁶ Mark Niese, “Georgia election data copied under direction of Trump attorney,” *The Atlanta Journal Constitution*, August 16, 2022. Available at: <https://www.ajc.com/politics/georgia-election-data-copied-under-direction-of-trump-attorney/XCPM33PXC5ABJN6UXG645EXLM4/>

¹⁷ Kate Brumback, “Video fills in details on alleged Ga. election system breach,” *The Associated Press*, September 6, 2022. Available at: <https://apnews.com/article/2022-midterm-elections-technology-donald-trump-voting-92c0ace71d7bee6151dd33938688371e>

logs.¹⁸ These show that the forensic images (containing the system software) have been uploaded over the Internet to an online ShareFile site and downloaded multiple times by other parties, including Conan Hayes. Hayes is implicated in the copying of voting system software in Mesa County, Colorado, and in the copying and unauthorized distribution of Antrim County, Michigan, software.¹⁹

Because credentials were shared and because the downloaded software may have been copied and shared, it is not possible to know how many people in the United States or abroad currently possess the Dominion Voting System software taken from Coffee County and used in hundreds²⁰ of other U.S. jurisdictions.

The *Curling* plaintiffs have amassed emails, contracts and other documentation that show SullivanStrickler was hired and paid by Sidney Powell to copy and distribute the software. The contracts show that Powell hired SullivanStrickler to access and copy voting system software in Georgia, Michigan, and Nevada, making this a multi-state enterprise.²¹

3. The Georgia state election officials' response has been inadequate, and doesn't reflect the severity of the breach, slowing state and federal law enforcement responses.

Evidence that the security of the voting system in Georgia may have been compromised first surfaced in December of 2020, when the Coffee County Election Director, Misty Hampton, posted a video to YouTube critical of the voting system.²² In the video, which went viral, the password for the election management server is visible on a post-it note on Hampton's computer. According to testimony from the Secretary of State's office, the office opened an investigation

¹⁸ See ECF No. 1518-1, available at <https://storage.courtlistener.com/recap/gov.uscourts.gand.240678/gov.uscourts.gand.240678.1518.1.pdf>; ECF No. 1518-2, available at <https://storage.courtlistener.com/recap/gov.uscourts.gand.240678/gov.uscourts.gand.240678.1518.2.pdf>.

¹⁹ See *supra* note 3.

²⁰ <https://verifiedvoting.org/verifier/#mode/search/year/2022/equipment/Ballot%20Marking%20Device/make/Dominion%20Voting%20Systems/model/ImageCast%20X%20BMD>

²¹ Emma Brown, Jon Swaine, Aaron C. Davis, Amy Gardner, "Trump-allied lawyers pursued voting machine data in multiple states, records reveal," *The Washington Post*, August 15, 2022. Available at: <https://www.washingtonpost.com/investigations/2022/08/15/sidney-powell-coffee-county-sullivan-strickler/>

²² Doug Richards, "Coffee County attention started with YouTube video," *11Alive*, September 29, 2022. Available at: <https://www.11alive.com/article/news/politics/elections/coffee-county-youtube-election-dominion-vote/85-14b18082-6f80-4657-8e37-b27e0d892735>

into the video and the posting of the password.²³ While the Secretary's investigators ultimately recommended penalties for failure to secure the doors to the election server room,²⁴ they apparently made no effort to review available video surveillance records to determine whether the server or server room security had been breached.²⁵ In fact, on January 26, 2021, the investigator charged with the investigation of the posted password and video happened upon one of the Cyber Ninja's colleagues, Jeffrey Lenberg, in Ms. Hampton's office with voting equipment made available to Lenberg for unauthorized access and testing.²⁶ There is no evidence that appropriate inquiry or further investigation ensued.

Misty Hampton resigned to avoid termination in February 2021.²⁷

In May 2021, her successor, James Barnes, wrote the Secretary of State's office because he was alarmed to find a business card from Doug Logan of Cyber Ninjas under the election director's computer.²⁸ Staff in the Secretary's office forwarded the message to the Secretary's chief investigator, asking her to investigate. According to an email produced in discovery, the Secretary's office planned to initiate an investigation to determine whether Cyber Ninjas had accessed the equipment.²⁹ The investigator acknowledged the email but, according to Barnes, no state investigation ensued.³⁰

Weeks later, Barnes needed to access the election management server, but found that the password he had been given did not work. He reported this to the Secretary's office. Technicians made a limited troubleshooting effort but were unable to access the server.³¹ They replaced the server and central scanner workstation on June 8, 2021.³² This incident, which suggested that someone had improperly changed the password, did not trigger any investigation into

²³ Germany Aff. ¶ 7 Document 1444-1 filed August 8, 2022.

²⁴ See State Election Board Transcripts for 2021, available at: https://sos.ga.gov/sites/default/files/2022-02/2021_seb.pdf

²⁵ Georgia Secretary of State Investigation Summary, September 28, 2021, available at: <https://www.dropbox.com/s/7jj34weu5pwy0a6/SEB%20investigation%2020-250.pdf?dl=0>

²⁶ "Georgia video, testimony at odds," *Northwest Arkansas Democrat Gazette*, September 21, 2022. Available at: <https://www.nwaonline.com/news/2022/sep/21/georgia-video-testimony-at-odds/>

²⁷ "Coffee Co. elections supervisor named in investigation resigns," *WALB News*, March 31, 2021. Available at: <https://www.walb.com/2021/03/31/coffee-co-elections-supervisor-named-investigation-resigns/>

²⁸ Barnes Dep. Page 55. Document 1440-1 filed July 29, 2022.

²⁹ Available at: [SOS investigation Doug Logan..pdf - Dropbox](#)

³⁰ Barnes Dep. Pages 160-168 Document 1440-1 filed July 29, 2022.

³¹ Barnes Dep. Pages 106-110 Document 1440-1 filed July 29, 2022.

³² Document 1377 and 1377-4.

unsanctioned voting system and software access in Coffee County. After taking possession of the potentially compromised server, the Secretary's office did not make meaningful efforts to access the contents of the server, nor did it examine the system forensically to determine whether there had been unauthorized access as Mr. Barnes had feared when he alerted the Secretary's office in May 2021.

In late February 2022, the plaintiffs shared with the Secretary's staff an audio recording of a call from an individual associated with the Trump campaign who was one of the organizers of the breach in Coffee County.³³ But in the following months, there was still no indication that the Secretary was investigating the allegations. On April 7, 2022, the Secretary of State's office represented to the court that it opened an investigation immediately after it received a recording of the phone call on March 2, 2022.³⁴ But records show the investigation was not actually opened until April 25, 2022—18 days *after* telling the court that the investigation had been opened.³⁵

Public statements from the Secretary's office suggest that for many months it failed to investigate the evidence of improper access to voting systems. In April 2022, Secretary Raffensperger claimed his office was investigating but found no evidence a breach had occurred.³⁶ That same month, at an event at the Carter Center, the Interim Georgia Deputy Secretary of State Gabriel Sterling summarily dismissed concerns of a breach in Coffee County, stating emphatically that the breach "did not happen."³⁷

While an investigation may have been "opened" on paper in late April 2022 by the Secretary and the State Election Board, their investigators made no efforts to interview Coffee County officials or potential witnesses, or request documents, obtain missing emails, or video surveillance records from Coffee County.³⁸ Nor was there a meaningful effort to forensically examine the seized server, now known to contain evidence of the January 7, 2021, breach.

³³ See *supra* note 2.

³⁴ April 7, 2022 hearing transcript, available at: <http://freespeechforpeople.org/wp-content/uploads/2022/11/court-transcript-4.7.22.pdf>

³⁵ Germany Aff. ¶ 26, Document 1444-1 filed August 8, 2022.

³⁶ See *supra* note 21.

³⁷ [Restoring Confidence in American Elections | Panel 3 \(April 29, 2022\) - YouTube](#) at 35:30

³⁸ Barnes Dep. Pages 160-161 Document 1440-1 filed July 29, 2022

Only in August 2022, months after the breach was publicly reported, did the Secretary—at the direction of the State Election Board Chair³⁹—bring in the Georgia Bureau of Investigation, which opened an investigation.⁴⁰

Thus, despite numerous concerning incidents, there is no indication Georgia state authorities did anything to investigate the statewide system breach until more than a year had passed. Sorting out the facts around the Coffee County incident and the Secretary’s slow response has been further complicated by conflicting statements provided by the Secretary to the press and courts.⁴¹ The Secretary’s delay in alerting Georgia law enforcement to this issue delayed Georgia authorities’ sharing information with the U.S. Department of Justice, Georgia Fusion Centers, and CISA—including the fact that this was a coordinated, multi-state plan orchestrated and funded by Donald Trump’s campaign attorneys.

4. Failure of the Georgia Secretary of State to respond to the security threats appropriately may also undermine a federal response to protect elections.

In public statements, the Georgia Secretary of State and his representatives have downplayed ongoing and escalating security threats stemming from the breach, claiming that the breach was sufficiently addressed by the removal of Misty Hampton and the replacement of the election server in 2021.⁴² The Secretary’s office has even dismissed concerns that the copying and distribution of the software poses ongoing risks as “fear-mongering.”⁴³

The inadequacy of the state’s response to this breach is clear in sworn testimony by Georgia’s own Chief Information Officer (CIO), Merritt Beaver.⁴⁴ In a deposition taken before the incident in Coffee County was confirmed, the CIO was asked hypothetically, “Can you explain what are the reasons that election

³⁹ Retired federal judge, The Honorable William S. Duffey, Jr., serves as the Chair of Georgia’s State Election Board.

⁴⁰ See *supra* note 3.

⁴¹ “Questions raised in timeline of state response to Coffee County breach,” *11Alive*, September 26, 2022. Available at: <https://www.11alive.com/article/news/politics/coffee-county-breach-timeline-georgia-secretary-of-state-brad-raffensperger-response/85-f3d75b6f-6ba8-445b-88d6-7fda894358be>

⁴² Gabriel Sterling on CNN’s “Out Front with Erin Burnett,” CNN Transcripts, September 7, 2022. Available at: <https://transcripts.cnn.com/show/ebo/date/2022-09-07/segment/01>

⁴³ <https://twitter.com/GabrielSterling/status/1576298738892836864>

⁴⁴ Mark Neisse, “Georgia election security chief splits time with second state job,” *Atlanta Journal Constitution*, August 5, 2022. Available at: <https://www.ajc.com/politics/georgia-election-security-chief-splits-time-with-second-state-job/SNR3WHSJIFB6RDAI3L6M7A43RE/>

software is not released to the public?” He responded, “It’s pretty obvious. You don’t expose your—basically your system to the public because they—basically you’re giving them a road map to how to basically get in and access the system.”⁴⁵ But the Secretary’s office has denied the unauthorized copying of the software poses any “long-term security threats,”⁴⁶ and has not adopted additional safeguards to protect future elections.

We raise these facts not to embarrass the Secretary’s office, but because they provide important context for federal agencies to respond efficiently and effectively. The failure of Georgia’s state election officials to investigate the events in Coffee County promptly, and the refusal of the Secretary to mitigate the danger to the entire statewide system demonstrate two things: 1) the Secretary’s office remains poorly informed about the events in Coffee County, and therefore the Department of Justice and the Department of Homeland Security should seek information from other sources, including from the plaintiffs and their attorneys in *Curling*; and 2) federal guidance is urgently needed to address the security risks posed by the unlawful access to and distribution of voting systems by entities that have demonstrated intentions to disrupt and destabilize U.S. elections *going forward*.

5. The Georgia system compromise is far more expansive and poses greater election security threats than the better understood serious breaches in other states.

Before the revelations of the breach of Georgia’s statewide system, county-level voting system breaches had been documented in multiple states including Michigan, Colorado, and Pennsylvania.⁴⁷ These serious security compromises are less threatening than the Georgia statewide breach, for multiple reasons:

⁴⁵ Beaver Dep. Document 1368-3 Page 157-158.

⁴⁶ Mark Niese, “Video shows fake elector aided copying of Georgia election data,” *The Atlanta Journal Constitution*, September 6, 2022. Available at: <https://www.ajc.com/politics/video-shows-fake-trump-electoral-aided-copying-of-georgia-election-data/NQM2F4KKMNGKRBHEAUSH6ALTGU/>

⁴⁷ Alexandra Ulmer and Nathan Layne, “Trump allies breach U.S. voting systems in search of 2020 fraud ‘evidence,’” *Reuters*, April 28, 2022. Available at: <https://www.reuters.com/investigates/special-report/usa-election-breaches/>

1) Unlike other states, Georgia uses the same software in each county: *every* Georgia county’s software was taken. The other states use multiple vendors and software versions, so a breach in any one county is not a statewide breach.

2) Unlike Michigan, Colorado, and Pennsylvania, the programming and configuration of the machines is coordinated at the state level, creating a “single point of failure” for an attacker who may want to attack multiple counties or polling locations.

3) Georgia is the only state using touchscreen Ballot-Marking Devices (“BMDs”) as the primary voting system for in-person voting. The paper trail produced by BMDs is inherently untrustworthy: it is a record of what the BMD did, not what the voters did. In contrast, the other states primarily use hand-marked paper ballots, which makes it possible for audits and recounts to detect and correct wrong outcomes. Georgia’s voting system does not support such auditing or error detection.⁴⁸

4) Georgia’s breach, unlike the others mentioned, encompassed the Election Management Server, scanners, touchscreen BMDs, electronic pollbooks, and all system components. Other states’ breaches generally obtained access to fewer voting-system components.

Another difference between the incidents in other states and in Coffee County, Georgia, is that everywhere else, the chief state election officials promptly engaged law enforcement, exercised their authority to contain the fallout, and made public the risk, enforcement measures, and mitigation efforts. State authorities investigated and enforced their laws and the cases were regarded as isolated. At the time, there was no public evidence that the breaches were connected—but now there is.

⁴⁸ Andrew Appel, Rich DeMillo, Philip Stark, “Ballot Marking Devices (BMDs) Cannot Ensure the Will of the Voters,” December 27, 2019. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3375755

6. Conclusion

One of the most significant revelations uncovered by the plaintiffs in *Curling v. Raffensperger* is that the Coffee County breach was just one element of a coordinated plan to access and copy voting system software from multiple states, multiple jurisdictions, and different voting system vendors,⁴⁹ by lawyers acting on behalf of the Trump campaign, possibly constituting federal crimes. Because this plot was orchestrated by individuals currently under investigation for their attempts to overturn the 2020 presidential election, it is possible that the coordinated effort to obtain voting system software was also part of an ongoing conspiracy to overturn elections. This calls for a vigorous and swift investigation by the Department of Justice, the Special Counsel, and the Federal Bureau of Investigation. Because the Georgia Secretary of State and State Election Board failed to investigate in a serious or timely manner, and to ensure that you get the most complete and accurate information regarding the Georgia breach, we urge you to seek information from the plaintiffs in *Curling*.

In addition, the Department of Homeland Security and Department of Justice's Division of Counterterrorism should immediately assess how the release of the software may affect the security of future elections and recommend security mitigations. Potential avenues of attack could include fabricating false evidence for disinformation campaigns, disrupting elections to prevent voting or cast doubt on the integrity of the process, or even altering the results. Your agencies' awareness, monitoring, and readiness for action is needed, since exploits may occur with little warning.

We thank you very much for your consideration and stand ready to assist in any way we can.

Sincerely,

Susan Greenhalgh
Senior Advisor for Election Security
Free Speech For People
Susan@FreeSpeechForPeople.org

Ron Fein
Legal Director
Free Speech For People
RFein@FreeSpeechForPeople.org

⁴⁹ See *supra* note 2.

Elizabeth Bradley Ph.D.
Professor
University of Colorado Boulder*

Duncan Buell Ph.D.
Chair Emeritus — NCR Chair in
Computer Science and Engineering
Dept. of Computer Science and
Engineering
University of South Carolina*

Lowell Finley
former Deputy Secretary of State
California

Richard A. DeMillo Ph.D.
Charlotte B. and Roger C. Warren
Professor of Computer Science
College of Computing
Georgia Institute of Technology*

Harri Hursti
Founding Partner Nordic Innovation
Labs
Election Integrity Foundation*

David Jefferson Ph.D.
Lawrence Livermore National
Laboratory* (retired)
Election Integrity Foundation*

Douglas W. Jones Ph.D.
Emeritus Associate Professor of
Computer Science, University of Iowa*

Daniel P. Lopresti Ph.D.
Professor, Department of Computer
Science and Engineering*
Chair, Computing Research
Association's Computing Community
Consortium (CCC)*
Past-President, International
Association for Pattern Recognition
(IAPR)*
Lehigh University

Peter G. Neumann Ph.D.
Chief Scientist,
SRI International Computer Science
Lab*

Mark Ritchie
Former MN Secretary of State
Member of the EAC Board of
Advisors*
Former president of the National
Association of Secretaries of State*

Philip B. Stark Ph.D.
Distinguished Professor
Department of Statistics
University of California, Berkeley*

John E. Savage Ph.D.
An Wang Professor Emeritus of
Computer Science
Brown University*

Penny M. Venetis
Director, International Human Rights
Clinic*
Distinguished Clinical Professor of
Law*
Judge Dickinson R. Debevoise
Scholar*
Center for Law and Justice
Newark, NJ

*Affiliations are listed for identification purposes only and do not imply institutional endorsement.



Increasing Concerns About Amplified Threats to Voting Systems

July 14, 2022 Palo Alto, CA. Over the past three weeks the OSET Institute has been encouraged to offer its position on questions about the public, but unauthorized availability of certain intellectual property (IP) assets related to certified, installed and operational voting systems and the potential risk or threat to election security involving those voting system technologies. It has been suggested that such a statement could be helpful to courts, legal counsels, media, and other interested members of the public. Ten percent of our public benefit services is election cyber-security advisory, which we have provided to state and federal agencies including elements of the national security apparatus for over a decade. The Institute's nonprofit mission also includes public education about election administration technology. Our senior most technologists have decades of experience in the analysis, design, and engineering of cyber, digital, and information security systems and services. And our team is fortified with several veteran election administrators. Backed by those capabilities and experience, we provide the following statement on the topic of risks posed by unauthorized public disclosure of election technology IP assets.

The Situation

Recent news about Dominion's voting system products has created concern that these products are at increased risk for compromise. Specifically,

1. CISA has issued a security advisory;
2. There was a recent wholesale disclosure of Dominion voting technology, related IP, and data in Coffee County, GA by improper access; and
3. Proceedings of a Georgia lawsuit could potentially result (at some point) in wider availability of information about Dominion products.¹

In the case of Coffee County, this included at least, but may not have been limited to, unauthorized acquisition and potential disclosure of disk images that contain all the data and software of an Election Management System (EMS). This is essentially the same type of information that was improperly obtained from Mesa County CO's elections office, and used as the basis for a quasi-technical report falsely claiming proof of a stolen election. In the Coffee County case, it remains unknown whether the acquisition and disclosure² of Dominion technology was less than, equal, or greater in breadth and depth, thereby exacerbating an already serious situation.

¹ To be clear, any disclosures here would be due to the Court releasing currently sealed documentation. There certainly is no intent to implicate any unauthorized disclosures or violations of protective orders by either party. Plaintiffs did perform a civic duty in tracking down and publicizing a breach by certain external non-governmental operatives abetted by insiders in Coffee County that occurred in 2020.

² It is also important to distinguish disclosure from publication; the first is bad enough, whereas the latter could be catastrophic. We are unaware of any publication in the Coffee County incident at this time, and it is unconfirmed how many have had access so far (i.e., recipients of a disclosure).

The Institute is seriously concerned about these unauthorized disclosures³ of election technology IP and the concerns outlined in our opening raise two questions:

1. Will the Coffee County (or any jurisdiction) release make it significantly easier for adversaries to attack future elections, or attack voting systems used in them?
2. If so, should these releases be stopped and future releases prevented (*if possible*)?

The short answer is that these releases are dangerous, regardless of any difference of opinion among technologists about whether such disclosure would be significantly easier, slightly easier, or simply catastrophic. And they must be prevented at all costs. Yet, their potential availability on the public Internet makes prevention a nearly impossible task— the proverbial cat is out of the bag.

Given the very challenging threat environment described next, even the slightest increase in ease of attack is an increase that must be avoided.

The Threat Environment

The current threat environment is daunting. Consider:

- The basic design of current voting system products (Dominion included) makes them fundamentally vulnerable to cyber-attack to change code, inject code, and/or tamper with data.
 - Voting system components are not fundamentally different from a typical personal computer (PC): subject to the installation of malware, to accidental or malicious modification of original software, modification of critical data by malicious or tampered software.
- Every independent review of every voting system in the last 15-years has shown that these products not only have these fundamental security weaknesses, but also are rife with poor quality software and many vulnerabilities, including those that are described in the dry specifics of CISA’s recent security advisory.
 - These findings apply not just to the voting machines that voters see, but also the election management system (EMS) that is the “back-office brains” of the entire voting system and its many parts.
- There are known flaws, vulnerabilities, exploits, and attackers.
 - Many known specific flaws, and many known attack vectors, including malware injection that can spread from EMS to voting machines or vice versa.⁴

³ <https://www.washingtonpost.com/investigations/2022/06/12/coffee-county-georgia-dominion-breach/>

⁴ The Election Management System (EMS) “sees” every record tagged with information that is unique to the jurisdiction. In the case of Dominion’s Democracy Suite EMS, it also handles ballot definitions, so it must both read and write those records. In addition to ballot definitions and voting records, the EMS also leaks considerable information about the development environment used to create the EMS software. If, for example, the developers’ network management suite used a version of SolarWinds with a certain release date (software subject to a massive attack recently), then an adversary would know the system might contain an exploitable back door. The same holds for the logging software. For our technical audience, one could, for instance, check to verify that unprintable “escape sequences” can be passed along to the server, which would thereby allow malicious software code to be injected when the EMS was operating to simply read and write records.

- High quality cyber attackers who are more than capable of using these attack vectors.
- Typical counter-measures, including those recommended by CISA, will not prevent attack, but raise physical and operational barriers that attackers have to circumvent—using known effective techniques—to gain access to target systems.
 - As voters, we cannot expect every single jurisdiction to properly and completely implement these countermeasures and to completely guard against attacks where attackers use access gained from insiders (usually inadvertently, e.g., via phishing)
- Disk images of EMS products are already known to be “in the wild,” (i.e., previously released on the so-called “dark web”) and available as a resource to attackers.
 - Advanced attackers do not need the information gained from these images, in order to construct successful attacks that alter election outcomes.
 - Security researchers have confirmed that the images contain information that can enable a larger set of adversaries to construct additional exploit techniques.⁵
 - However, broader disclosure and distribution of deeper amounts of election technology IP—such as what was experienced in Coffee County—greatly expands the scope of opportunity for development of more attack methods and mechanisms, and we believe this is game-changing for the threat environment.

The Consequences of Disclosures

Given this threat environment, it must be unacceptable to release such vendor IP, including, but not limited to disk images, because they provide a powerful increment of information to broaden the range of exploits on voting systems used in elections. The recent unauthorized disclosure and the apparent quantity and quality of software, data, and other artifacts could trigger a tsunami of new attack methods, means, and mechanisms. In fact, these disclosures (and worse, publication) provide massive updates to materials already “in the wild” and represent a clear and present danger to election infrastructure protection and security.

⁵ Imaging generally means a binary or executable image that may look intimidating to the untrained eye, but is actually the entry point for significant vulnerability analysis. While not immediately “readable” for understanding operational capabilities, the binary or executable code can be parsed and read by readily available reverse engineering tools (including disassemblers and reverse compilers). In fact, such tools are regularly used in introductory computer science college courses. We also note that many commentators and recreational programmers tend to focus on software source code (which suggests reverse compilation); however, disassemblers are more useful for analysis because they assign English language tags and commands. Those are used to answer the analysts’ questions (bear in mind that vulnerability discovery is a series of experiments—the analyst asks whether the system checks some value and then hunts for the code that answers that particular question).

In practice, and to the casual observer, this may appear to be poking around in the darkness of an incomprehensible string of binary code (1s and 0s). However, in fact, once that process is underway and one begins to identify APIs, libraries, and hard-coded information, then standard tests can be applied (again, as in lab work in every computer science curriculum) and the process of discovery and enlightenment accelerates.

However, there is another related concern. An additional, more recent, factor is that some election administration or operations insiders appear to be at least abetting, and at worst, performing abuse of voting systems for some (nefarious and/or partisan) agenda. As a result:

- More access to disk images and disclosed (or worse, published) content means more opportunity to develop new exploits in addition to those that do not require such access.
- More exploits become available, including exploits that can be used by larger numbers of less sophisticated attackers operating locally.
- An increased pool of exploits and attackers, when combined with insider access (i.e., physical infrastructure security is compromised), creates many more opportunities for effective attacks.

Returning to the original questions, it's not necessary to assess the increased risk or likelihood of an actual attack, or to assess the likelihood of a larger pool of attackers or insiders. It is simply the possibility of enabling a larger number of attackers with a larger array of exploits, that should be enough to take measures to prevent unauthorized release of disk images, data dumps, or other technical internals of voting system products that have been shown to be rife with security issues.

Conclusions

Further unauthorized release and publication of election technology IP must be prevented because they can, and likely will facilitate and foster future attacks on election trust through either cyber-weapons or disinformation weapons.

Assessing the *likelihood* of such attacks or their impact is not required; the threat environment is so daunting, and public trust in U.S. elections is so fragile, that any amplification of any nature must be prevented. In any event, once these disclosures have occurred, significant efforts and resources must be deployed to immediately mitigate the misinformation likely to be created and amplified to further deteriorate public confidence.

Finally, unauthorized disclosures that lead to such amplifications must not be tolerated, and those responsible held accountable to the full extent of the law. And a related topic that emerges for consideration elsewhere is strengthening laws, regulations, and penalties for such compromises of what amounts to national security assets given election technology is critical infrastructure.

In closing, we also find it important to add the following clarification. None of the content in this statement about the dangers of unauthorized intellectual property releases (e.g., disk images⁶ et al) is meant to suggest, support, or make any case for any claim that the 2020 election may have been tampered with, altered, hacked, or rigged. Such is absolutely not the case because there is no evidence whatsoever that any such digital subversive act occurred in any system in any of the jurisdictions where such claims were made.

⁶ We note in passing that disk imaging is a detailed examination—an intimate act of inspection. The very existence of an image file discloses much about the physical security or lack thereof in the vendor's development environment. The OSET Institute is a strong proponent of high assurance engineering for the design and development of the critical infrastructure of election administration technology. Disk imaging is a good example of why.

Moreover, for Georgia, the durable paper ballot of record remains available and was used to verify the results through recounting. However, that process itself has not been without its own complications.⁷

We also observe that none of this statement about the unauthorized disclosure of election technology IP has anything to do with the court-sealed report by University of Michigan's Dr. Halderman subject of pending GA litigation, which addresses separate vulnerabilities in other parts of the Dominion election system.

#####

⁷ We do not mean to suggest that, in the case of Georgia, the Ballot Marking Device (BMD) printout is necessarily trustworthy and or that there are routine risk-limiting audits or recounts of all contests everywhere. That's simply not the case. Our point here is that in the face of unsupported claims about subversive manipulation of voting machines in 2020, there was other physical evidence to consider in order to verify the results.

And the over-arching point is that unauthorized acquisition, disclosure, or publication of voting system vendors' intellectual property in the form of disk images or any other means in order to attempt to prove a theory about election hacking is absolutely not the way to do so.

Further, to the nuance of Georgia ballot counting and recounting, one of the main issues in the *Curling v. Raffensperger* case as we understand it, is that the BMDs might not print voters' selections accurately, either on the plaintext or the QR code. Moreover, Georgia only audits one contest every 2-years, and candidly we are not convinced their audits are rigorously conducted. Yet, we need not single out Georgia on that count; in fact, most states do not perform post-election results audits in a way that — even if the paper trail were demonstrably trustworthy — has any chance of correcting wrong outcomes, much less a guaranteed minimum chance (*limiting risk*). And to that end, Georgia does not appear to rigorously keep track of the paper (or memory cards, etc.), which is aggravated by the fact that in Georgia the QR codes are the legal official vote for counts and recounts. So, yes, paper ballots are the very best way to start toward a post-election verification of the outcome and an excellent means to thwart “kraken” about subversive activities in the machinery. However, the reliability of the paper cannot be presumed absent certain processes, procedures, and clear evidence of voter intent in the ballot as cast.

Freedom to Tinker

Research and commentary on digital technologies in public life

Magical thinking about Ballot-Marking-Device contingency plans

JULY 21, 2022 BY ANDREW APPEL LEAVE A COMMENT

The Center for Democracy and Technology recently published a report, “**No Simple Answers: A Primer on Ballot Marking Device Security**”, by William T. Adler. Overall, it’s well-informed, clearly presents the problems as of 2022, and it’s definitely worth reading. After explaining the issues and controversies, the report presents recommendations, most of which make a lot of sense, and indeed the states should act upon them. **But there’s one key recommendation in which Dr. Adler tries to provide a simple answer, and unfortunately his answer invokes a bit of magical thinking. This seriously compromises the conclusions of his report. By asking but not answering the question of “what should an election official do if there are reports of BMDs printing wrong votes?”, Dr. Adler avoids having to make the inevitable conclusion that BMDs-for-all-voters is a hopelessly flawed, insecurable method of voting. Because the answer to that question is, unfortunately, there’s *nothing* that election officials could usefully do in that case.**

BMDs (ballot marking devices) are used now in several states and there is a serious problem with them (as the report explains): “a hacked BMD could corrupt voter selections systematically, such that a candidate favored by the hacker is more likely to win.” That is, if a state’s BMDs are hacked by someone who wants to change the result of an election, the BMDs can print ballots with votes on them different from what the voters indicated on the touchscreen. Because most voters won’t inspect the ballot paper carefully enough before casting their ballot, most voters won’t notice that their vote has been changed. The voters who *do* notice are (generally) allowed to “spoil” their ballot and cast a new one; but the substantial majority of voters, those who don’t check their ballot paper carefully, are vulnerable to having their votes stolen.

One simple answer is not to use BMDs at all: let voters mark their optical-scan paper ballots with a pen (that is, HMPB: hand-marked paper ballots). A problem with this simple answer (as the report explains) is that some voters with disabilities cannot mark a paper ballot with a pen. And (as the report explains) if BMDs are reserved just for the use of voters with disabilities, then those BMDs become “second class”: pollworkers are unfamiliar with how to set them up, rarely used machines may not work in the polling place when turned on, paper ballots cast by the disabled are distinguishable from those filled in with a pen, and so on.

So Dr. Adler seems to accept that BMDs, with their serious vulnerabilities, are inevitably going to be adopted—and so he makes recommendations to mitigate their insecurities. And most of his recommendations are spot-on: incorporate the cybersecurity measures required by the VVSG 2.0, avoid the use of bar codes and QR codes, adopt risk-limiting audits (RLAs). Definitely worth doing those things, if election officials insist on adopting this seriously flawed technology in the first place.

But then he makes a recommendation intended to address the problem that if the BMD is cheating then it can print fraudulent votes that will survive any recount or audit. The report recommends,

Another way is to depend on voter reports. In an election with compromised BMDs modifying votes in a way visible to voters who actively verify and observe those modifications, it is likely that election officials would receive an elevated number of reported errors. In order to notice a widespread issue, election officials must be monitoring election errors in real-time across a county or state. If serious problems are revealed with the BMDs that cast doubt on whether votes were recorded properly, either via parallel testing or from voter reports, election officials must respond. **Accordingly, election officials should have a contingency plan in the event that BMDs appear to be having widespread issues.** Such a plan would include, for instance, having the ability to substitute paper ballots for BMDs, decommissioning suspicious BMDs, and investigating whether other machines are also misbehaving. Stark (2019) has warned, however, that because it is likely not possible to know how many or which ballots were affected, the only remedy to this situation may be to hold a new election.

This the magical thinking: “election officials should have a contingency plan.” The problem is, when you try to write down such a plan, there’s nothing that actually works! Suppose the election officials rely on voter reports (or on the rate of spoiled ballots); suppose the “contingency plan” says (for example) says “if x percent of the voters report malfunctioning BMDs, or y percent of voters spoil their ballots, then we will . . .” Then we will what? Remove those BMDs from service in the middle of the day? But then all the votes *already cast* on those BMDs will have been affected by the hack; that could be thousands of votes. Or what else? Discard all the paper ballots that were cast on those BMDs? Clearly you can’t do that without holding an entirely new election. And what if those x% or y% of voters were *fraudulently* reporting BMD malfunction or *fraudulently* spoiling their ballots to trigger the contingency plan? There’s no plan that actually works.

Everything I’ve explained here was already written down in “**Ballot-marking devices cannot ensure the will of the voters**” (2020 [**non-paywall version**]) and in “**There is no reliable way to detect hacked ballot-marking devices**” (2019), both of which Dr. Adler cites. But an important purpose of magical thinking is to avoid facing difficult facts.

It’s like saying, “to prevent climate change we should just use machines to pull 40 billion tons of CO₂ out of the atmosphere each year.” But there is no known technology that can do this. All the direct-air-capture facilities deployed to date **can capture just 0.00001 billion tons**. Just because we *really, really want* something to work is not enough.

There is an inherent problem with BMDs: they can change votes in a way that will survive any audit or recount. Not only is there “no simple solution” to this problem, there’s no solution period. Perhaps someday a solution will be identified. Until then, BMDs-for-all-voters is dangerous, even with all known mitigations.

FILED UNDER: [UNCATEGORIZED](#) TAGGED WITH: [VOTING](#)

Speak Your Mind

Name *

Email *

Post Comment

Freedom to Tinker is hosted by Princeton’s Center for Information Technology Policy, a research center that studies digital technologies in public life. Here you’ll find comment and analysis from the digital frontier, written by the Center’s faculty, students, and friends.



Search this website ...

Search

- What We Discuss
- [AACs](#) [bitcoin](#) [CD](#) [Copy Protection](#) [censorship](#) [CITP](#) [Competition](#) [Computing in the Cloud](#) [Copyright](#) [Cross-Border Issues](#) [cybersecurity](#) [policy](#) [DMCA](#) [DRM](#) [Education](#) [ethics](#) [Events](#) [Facebook](#) [FCC](#) [Government](#) [Government transparency](#) [Grokster](#) [Case](#) [Humor](#) [Innovation Policy](#) [Internet](#) [Law](#) [Managing the](#) [Internet](#) [Media](#) [NSA](#) [Online](#) [Communities](#) [Peer-to-Peer](#) [Predictions](#) [Princeton](#) [Privacy](#) [Publishing](#) [Recommended](#) [Reading](#) [Secrecy](#) [Security](#) [Spam](#) [Super-](#) [DMCA](#) [surveillance](#) [Tech/Law/Policy](#) [Blogs](#) [Technology and](#) [Freedom](#) [Voting](#) [transparency](#) [Wiretapping](#) [WPM](#)

Contributors

Select Author...

- Archives by Month
- 2022: J F M A M J J A S O N D
 - 2021: J F M A M J J A S O N D
 - 2020: J F M A M J J A S O N D
 - 2019: J F M A M J J A S O N D
 - 2018: J F M A M J J A S O N D
 - 2017: J F M A M J J A S O N D
 - 2016: J F M A M J J A S O N D
 - 2015: J F M A M J J A S O N D
 - 2014: J F M A M J J A S O N D
 - 2013: J F M A M J J A S O N D
 - 2012: J F M A M J J A S O N D
 - 2011: J F M A M J J A S O N D
 - 2010: J F M A M J J A S O N D
 - 2009: J F M A M J J A S O N D
 - 2008: J F M A M J J A S O N D
 - 2007: J F M A M J J A S O N D
 - 2006: J F M A M J J A S O N D
 - 2005: J F M A M J J A S O N D
 - 2004: J F M A M J J A S O N D
 - 2003: J F M A M J J A S O N D
 - 2002: J F M A M J J A S O N D

author log in



DVSorder

Mitigation Tool

Technical Details

Published October 14, 2022

DVSorder is a privacy flaw that affects Dominion Voting Systems (DVS) ImageCast Precinct (ICP) and ImageCast Evolution (ICE) ballot scanners, which are used in parts of [21 states](#). Under some circumstances, the flaw could allow members of the public to identify other peoples' ballots and learn how they voted.

This vulnerability is a privacy flaw and *cannot* directly modify results or change votes. Nevertheless, the secret ballot is an important security mechanism, and some voters—especially the most vulnerable in society—may face real or perceived threats of coercion unless the privacy of their votes is strongly protected.

Many jurisdictions [publish data](#) from individual voted ballots, such as cast-vote records (the votes from each ballot) or ballot images (scans of each ballot). This data is usually supposed to be randomly shuffled, to protect voters' privacy. The DVSorder vulnerability makes it possible to [unshuffle the ballots](#) and learn the order they were cast. This sometimes [makes it possible](#) to determine how specific individuals voted.

Jurisdictions can continue to publish ballot-level data if they take steps to “sanitize” data from vulnerable Dominion scanners. We have created a [sanitization tool](#) to help. Public access to election data, including cast-vote records and ballot images, can be valuable for voter confidence, and DVSorder is not a reason to reduce transparency.

We were [able to discover](#) the vulnerability using only publicly available information, and it could potentially be discovered and exploited by anyone, without any access to equipment or breach of controls. Although sanitizing data will protect against exploitation by the public, the original copies of the records remain vulnerable. This means there will still be risks from insiders or data breaches until the scanners are eventually [patched](#). We are [making our findings public](#) to ensure all localities are

informed in time to avoid releasing vulnerable data from the November election. We [alerted](#) Dominion, CISA, EAC, and state officials prior to publication.

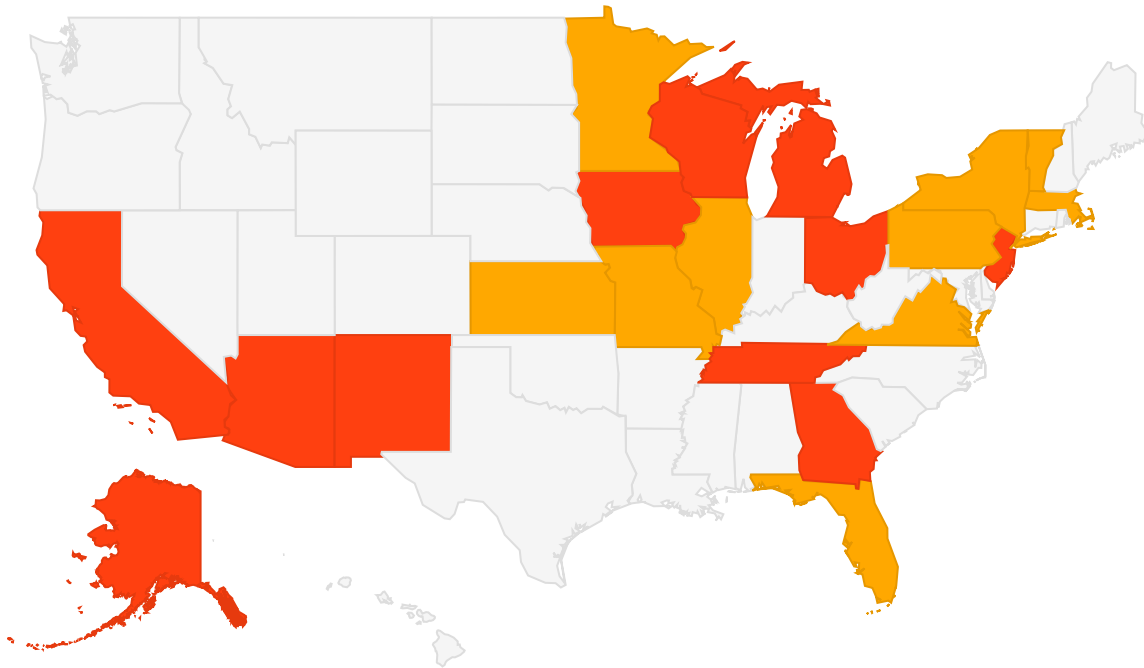
This research was conducted by [Braden Crimmins](#), Dhanya Narayanan, Josiah Walker, and [J. Alex Halderman](#) at the University of Michigan and [Drew Springall](#) at Auburn University. We can be contacted at team@DVSorder.org.

- [Which jurisdictions are at risk?](#)
- [How does the vulnerability work?](#)
- [How does knowing the ballot order threaten privacy?](#)
- [What machines and kinds of data are vulnerable?](#)
- [How can election officials mitigate this?](#)
- [Is there a software patch?](#)
- [What disclosure was made prior to publication?](#)
- [Why are you publishing this before the election?](#)
- [What are the technical details?](#)



Which jurisdictions are at risk?

According to data from [Verified Voting](#), parts of 21 states and Puerto Rico use the vulnerable Dominion scanners. So far we have identified jurisdictions in 11 states that appear to have published vulnerable data from recent elections: Alaska, Arizona, California, Georgia, Iowa, Michigan, New Jersey, New Mexico, Ohio, Tennessee, and Wisconsin.



Some jurisdictions make this shuffled ballot data public, most commonly in the form of *ballot images* (scans of each individual ballot) or *cast vote records* (data files that record the votes from individual ballots). [Dominion's documentation](#) implies that the shuffled data can be safely distributed without compromising voters' privacy, as does [information Dominion provided](#) during state equipment purchasing: "*The ballot images are given a random ID number as their file name, and when the images are extracted by the [EMS] application, they are randomized, thus ensuring the ballot images are de-coupled from voter order.*"

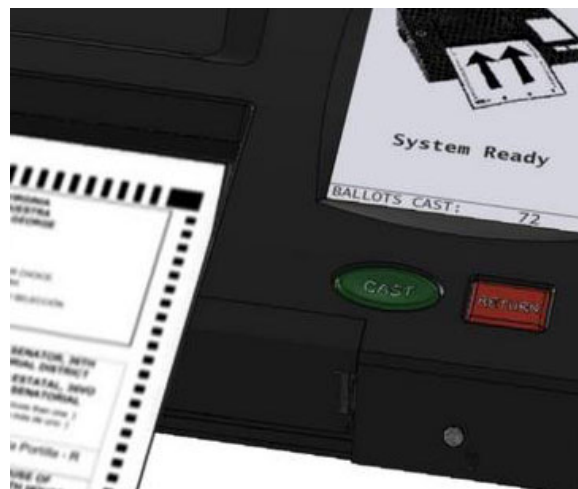
Unfortunately, the Dominion ICP and ICE scanner software is flawed such that ballot record IDs are assigned in a [predictable manner](#). This allows anyone to unshuffle the ballot images or cast vote records and learn the order in which they were cast.

Although the DVSSorder vulnerability is specific to two models of Dominion scanners, we recommend that other voting equipment vendors review the [technical details](#) and confirm that their implementations do not reveal the order in which ballots were cast.

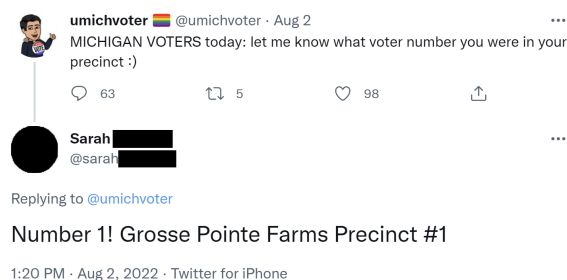
How does knowing the ballot order threaten privacy?

There are several types of scenarios where the DVSSorder vulnerability could be exploited to identify how specific people voted:

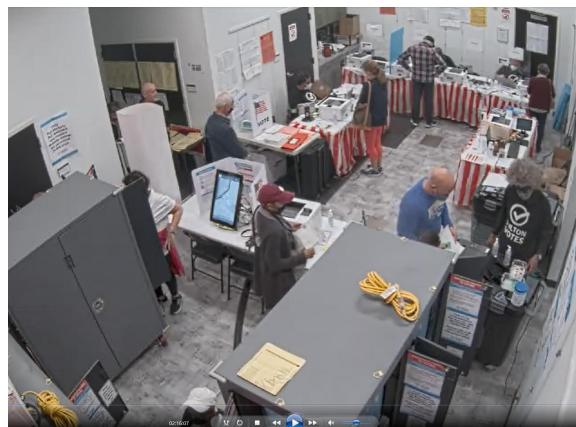
- In most jurisdictions, scanners display a public counter that shows how many ballots have been cast. Anyone can note the counter value when they vote and thereby learn the ballot sequence numbers of people who vote before and after. For example, suppose a man uses the scanner immediately after his wife. By noting the counter value just before scanning his ballot, the man can later identify his wife's ballot in published cast vote records or ballot image data and see how she voted.



- Poll workers or election observers could similarly note the public counter value to target specific voters. They could also keep a complete record of who uses the scanner, in order, which would allow them to deanonymize all ballots cast at the precinct.
- Some voters publicly disclose their polling places and voter numbers on social media or to others, as in the tweet shown here. As long as the voter has accurately stated their position in the ballot sequence, this would allow anyone to determine how they voted from vulnerable CVRs or ballot images.



- Some localities record all-day surveillance footage inside polling places. (This image is from a day-long video from a county in Georgia and was obtained by others prior to our work via a public records request.) If the jurisdiction releases vulnerable CVRs or ballot images, anyone could associate each ballot with footage of the voter casting it. A larger number of jurisdictions treat voter check-in records or poll books as public records. These can heighten the risks posed by the vulnerability, as they often track the order in which voters receive their ballots, which can match or closely approximate the order of casting.
- Some localities publish scanner log files (`slog.txt` files) from the ICP or ICE. Although these logs by themselves pose little risk to privacy, they can be combined with the DVSError vulnerability to determine the *exact time* that each CVR or ballot image was cast (subject to the accuracy of the scanner's internal clock). This provides an additional route to identify voters' ballots. As one example, journalists sometimes film or photograph candidates and



other political figures as they vote. Such media is often timestamped and could be used by anyone to deanonymize those individuals' ballots, even long after the election.

What machines and kinds of data are vulnerable?

All versions of the Dominion ICP and ICE for which we have located public ballot-level data appear to be vulnerable to DVSorter, including versions that have been [certified](#) by the U.S. Election Assistance Commission (EAC). The problem is specific to the ICP and ICE; ImageCast Central scanners and ImageCast X DREs do not appear to suffer from the flaw. (The ImageCast Central (ICC) intentionally labels ballots in the order they are scanned.)



ImageCast Precinct (ICP)



ImageCast Evolution (ICE)

Dominion's EMS software can export ballot-level data in several forms. Some examples of the most commonly published types of data that may be vulnerable are:

```
"TabulatorId": 145062,
"BatchId": 0,
"RecordId": 180517,
"CountingGroupId": 2,
"ImageMask": "D:\\NAS\\20201103 General\\Results\\
\\Images\\145062_00000_180517*.\"",
"SessionType": "ScannedVote",
"VotingSessionIdentifier": "",
"UniqueVotingIdentifier": "",
"Original": {
  "PrecinctPortionId": 1166,
  "BallotTypeId": 1076,
```

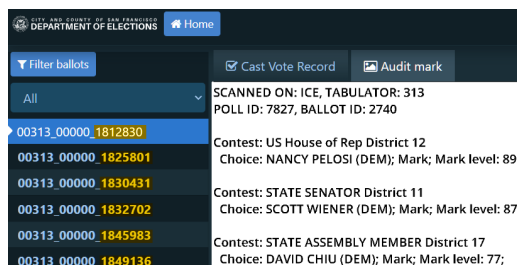
JSON cast-vote records (CVRs)

TabulatorNum	BatchId	RecordId	CountingGroup	PrecinctPortion
987	0	546984	Election Day	City of Detroit, P
987	0	749347	Election Day	City of Detroit, P
987	0	299063	Election Day	City of Detroit, P
987	0	607574	Election Day	City of Detroit, P
987	0	964848	Election Day	City of Detroit, P
987	0	60943	Election Day	City of Detroit, P
987	0	882503	Election Day	City of Detroit, P
987	0	317204	Election Day	City of Detroit, P
987	0	989938	Election Day	City of Detroit, P
987	0	700677	Election Day	City of Detroit, P

CSV cast-vote records (CVRs)



Ballot image TIF files
(with record IDs in filenames)



Ballot audit website
(with record IDs in filenames)

Only data that represents individual ballots and their record IDs is affected by DVSSorder. Summary results such as statements of votes cast (SoVCs), precinct- or scanner-level totals, election-night result reports, and poll tapes are not vulnerable to this privacy flaw.

How can election officials mitigate this?

DVSSorder affects only two specific models of ballot scanners: the Dominion ImageCast Precinct (ICP) and ImageCast Evolution (ICE). Jurisdictions that do not use these scanners are unaffected and do not need to take any action. DVSSorder should not motivate unaffected jurisdictions to decrease their transparency.

Localities that use the Dominion ICP or ICE can prevent the flaw from being exploited by the public by taking specific steps to “sanitize” ballot-level data before publishing it:

Manually Sanitizing CVRs (CSV format only)

Dominion cast-vote records (CVRs) in CSV format use a simple data scheme that can be sanitized manually. To do so, open the .csv file in Excel and [delete column D](#), labeled “RecordId”, then save the file. Removing the record IDs from JSON-format CVRs and ballot image filenames is more labor intensive, so we recommend using our data sanitization tool described below.

Automated Data Sanitization Tool (all formats)

We created an open-source software tool that can automatically reprocess Dominion cast-vote records (CVRs) and ballot image files so that DVSSorder can no longer be exploited. The tool can sanitize CVRs in .csv or .zip format and folders of ballots images in .tif format.

Sanitizing published ballot-level data cannot affect official election results, because results are generated directly from the election management system (EMS), not from the ballot-level data released to the public. However, as with any third-party software,

jurisdictions *should not* run our sanitization tool on their EMS computers. Instead, we recommend copying vulnerable CVRs or ballot images to an external system and running the tool there. Our tool is open-source software, and we encourage anyone interested to view the code and test its behavior.

More about our tool:

- [Read the documentation](#)
- [View the source code](#)

Election officials who need assistance can [contact us](#), and we will be happy to provide any help we can.

Is there a software patch?

[Sanitizing ballot-level data](#) before publishing it makes the data just as safe to release as if the DVSorder vulnerability did not exist. However, even if jurisdictions sanitize the data they make public (or if they do not publish any ballot-level data), the flaw still carries risks. For instance, unsanitized data could be stolen in a data breach or accessed by malicious insiders, who could exploit the flaw to learn how people voted.

Completely mitigating these risks will require Dominion to change the ICP and ICE firmware to use a secure method of generating ballot IDs. The U.S. Election Assistance Commission (EAC) has informed us that Dominion plans to correct the flaw in future firmware versions. However, our understanding is that no patches will be available until after the November election, at least for federally certified versions of Dominion systems. Election officials should contact Dominion for further information and to inquire as to patch availability.

What disclosure was made prior to publication?

We notified Dominion about the vulnerability on August 23, 2022. Our [disclosure letter to Dominion](#) informed them that we planned to publish information about the flaw as soon as 30 days later and offered to assist them in understanding and mitigating the problem. The company acknowledged receipt of the disclosure on August 29, but we have not received any subsequent communication from them. We informed the U.S. Election Assistance Commission (EAC) and the Cybersecurity & Infrastructure Security Agency (CISA) about the vulnerability on September 2.

Two weeks after we notified Dominion, it sent a “[customer notification](#)” to jurisdictions that use the ICP and ICE. (Dominion did not provide us a copy, but we

obtained one from an affected jurisdiction.) While the notification appears to be in response to our disclosure, it **does not mention** that the scanners have a vulnerability that reveals the order in which ballots were cast. Instead, it directs election officials to *“follow any state or local requirements guiding public access to and release of cast vote records”* and to *“consult their legal advisors for guidance on how best to ensure that [voter secrecy] protections are applied, particularly if simultaneously releasing any record (i.e. [sic] video) that could reveal a voter’s identity in the order in which they cast their ballot.”*

We observe that such legal advisors would likely rely on Dominion’s prior, inaccurate representations that ballot-level data is appropriately randomized to protect privacy. By failing to provide information about the specific risks posted by the DVSorder flaw, Dominion’s notice appears to have left jurisdictions unable to make informed decisions about whether and how to release election data.

Before publication of this website, we sent our own notifications to the state election directors in states that we believe use ICP or ICE scanners.

Why are you publishing this before the election?

We consulted with other experts and considered a range of equities before concluding that the public interest would be best served by publishing now.

The vulnerability is unusual in that it doesn't require exotic skills or special access to discover or exploit, but rather only publicly available information. This means there is an appreciable risk that malicious parties would independently discover the flaw, or that they already have. With the bar so low, we're concerned that people will attempt to exploit it following the midterms.

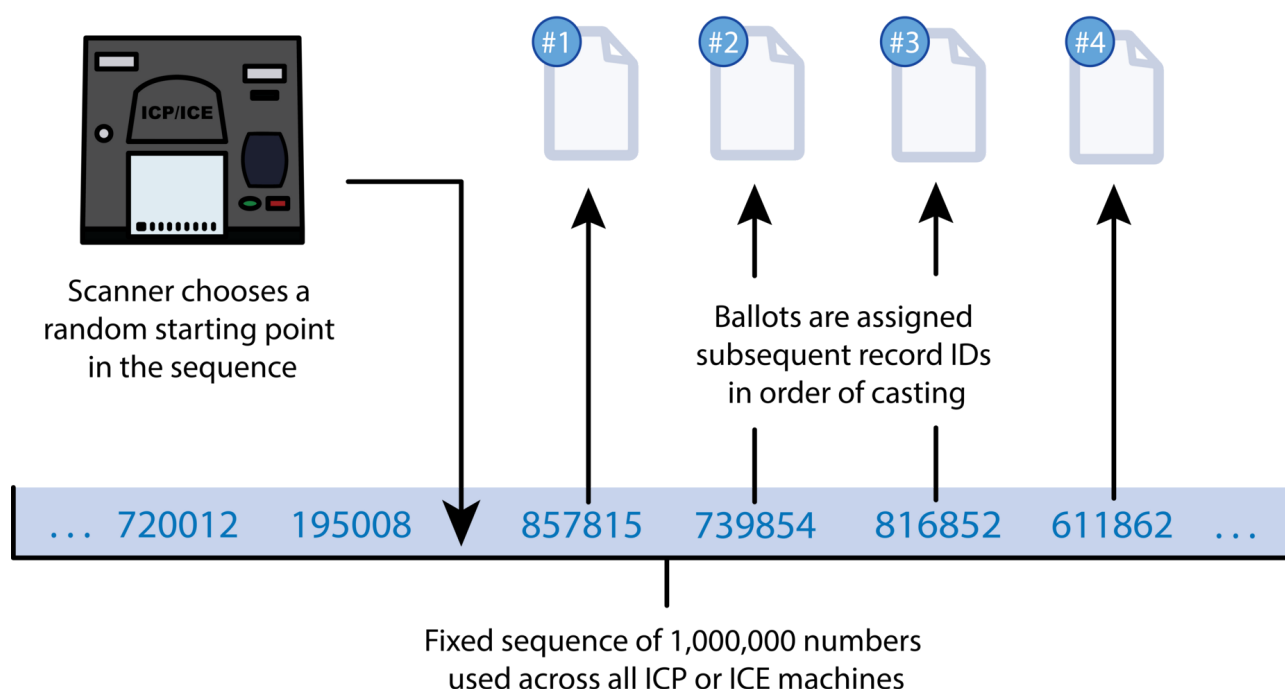
If we did not make our findings public before the election, jurisdictions would almost certainly publish a large volume of vulnerable data in November. Once released, this data would remain vulnerable in perpetuity, even if the scanners themselves were later patched. Raising the alert now gives election officials time to respond effectively. Our priority is to prevent this flaw from affecting voters in the midterms, which is ultimately the best way to uphold public trust.

Technical details

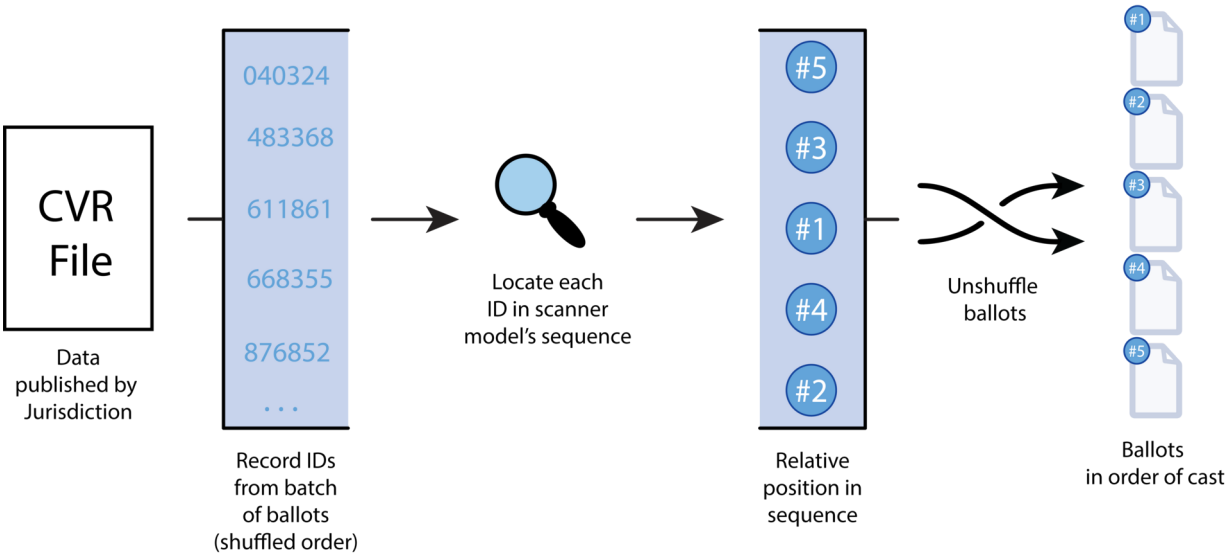
The Dominion ICP and ICE generate ballot record IDs using a pseudorandom number generator (PRNG). The PRNG they use is based on a [linear congruential generator](#)

(LCG). LCGs have long been known to be unsuitable for most security applications, both because the sequences they generate have obvious patterns and because their entire output is predictable given only a few samples. Dominion attempts to obfuscate the LCG output using some simple transformations (which differ slightly between the ICP and ICE), but these are insufficient to make the PRNG secure.

The ICP and ICE PRNGs each output a fixed sequence of 1,000,000 numbers (a permutation of the numbers 0-999,999) that is the same across all devices of each model. For a given batch of ballots, only the *starting point* within the sequence is randomized. The ballot record IDs are simply consecutive values in the fixed sequence from that point forward:



If an attacker knows the record IDs from the ballots in a batch (from CVRs, ballot image filenames, or any other source), they merely need to locate them in the PRNG output sequence for the scanner model. The record ID that appears first in the sequence corresponds to the earliest ballot, and all other record IDs will appear following it in the sequence, in the order in which they were cast:



We identified the vulnerability from just a short series of record IDs in voted order, which we obtained from publicly available data. Even in a small sample (like the example shown below), there are clear repeating patterns in several of the digit positions. This immediately suggests the use of a simple, non-cryptographic PRNG, such as an LCG:

303001	907271	801991
720012	224222	722982
195008	599278	693998
857815	956625	858485
739854	332644	435414
611861	513631	617401
876852	170642	074412
483368	385138	481708
668355	764145	266715
040324	149184	042754

Starting from this observation, multiple members of the team were able to independently reconstruct the complete PRNG algorithm within a few days.

Both the ICP and ICE PRNGs generate record IDs through a simple sequence of steps. They start with the LCG $x_{n+1} = x_n + 864,803 \bmod 1,000,000$. The output is then obfuscated by a simple substitution cipher in which the digits [0,1,2,3,4,5,6,7,8,9] are replaced by [5,0,8,3,2,6,1,9,4,7]. The digits are then reordered following a fixed permutation that is different on the ICP and the ICE.

This code reproduces the complete record ID sequence for each scanner model:


```

1 def generate_sequence(p):
2     return [sum([5,0,8,3,2,6,1,9,4,7][864803*n//10**p[i]%10]*10**i for i in range(6
3         for n in range(1000000))]
4 icp_sequence = generate_sequence([2,3,1,5,0,4])
5 ice_sequence = generate_sequence([1,5,0,4,2,3])

```

dvs_prng.py hosted with ❤ by GitHub

[view raw](#)

To validate and test for the vulnerability, we created a [proof-of-concept implementation](#). This program inputs a CSV- or JSON-format CVR file and outputs the fraction of ballots that appear to be vulnerable.

We will provide additional technical details in a forthcoming research paper.

The DVSorder website and logo ([svg](#); free to use under a [CC0](#) license) were designed by [Sarah Madden](#). Technical illustrations are by LaKyla Thomas.



Georgia
Secretary of State
Brad Raffensperger

- SOS Office ▾
- Business ▾
- Charities ▾
- Elections ▾
- Securities ▾
- Licensing ▾
- Search ▾

Historic First Statewide Audit of Paper Ballots Upholds Result of Presidential Race

[Home](#) > [News & Announcements](#) > Historic First Statewide Audit of Paper Ballots Upholds Result of Presidential Race

November 19th, 2020

(Atlanta) -- Today, Secretary of State Brad Raffensperger announced the results of the Risk Limiting Audit of Georgia’s presidential contest, which upheld and reaffirmed the original outcome produced by the machine tally of votes cast. Due to the tight margin of the race and the principles of risk-limiting audits, this audit was a full manual tally of all votes cast. The audit confirmed that the original machine count accurately portrayed the winner of the election. The results of the audit can be viewed [HERE](#) , [HERE](#) , and [HERE](#) .

“Georgia’s historic first statewide audit reaffirmed that the state’s new secure paper ballot voting system accurately counted and reported results,” said Secretary Raffensperger. “This is a credit to the hard work of our county and local elections officials who moved quickly to undertake and complete such a momentous task in a short period of time.”

“Georgia’s first statewide audit successfully confirmed the winner of the chosen contest and should give voters increased confidence in the results,” said Ben Adida, Executive Director of VotingWorks. “We were proud to work with Georgia on this historic audit. The difference between the reported results and the full manual tally is well within the expected error rate of hand-counting ballots, and the audit was a success.”

By law, Georgia was required to conduct a Risk Limiting Audit of a statewide race following the November elections. Understanding the importance of clear and reliable results for such an important contest, Secretary Raffensperger selected the presidential race in Georgia for the audit. Meeting the confidence threshold required by law for the audit meant conducting a full manual tally of every ballot cast in Georgia.

The Risk Limiting Audit reaffirmed the outcome of the presidential race in Georgia as originally reported, with Joe Biden leading President Donald Trump in the state.

The audit process also led to counties catching making mistakes they made in their original count by not uploading all memory cards. Those counties uploaded the memory cards and re-certified their results, leading to increased accuracy in the results the state will certify.

The differential of the audit results from the original machine counted results is well within the expected margin of human error that occurs when hand-counting ballots. A 2012 [study](#) by Rice University and Clemson University found that “hand counting of votes in postelection audit or recount procedures can result in error rates of up to 2 percent.” In Georgia’s recount, the highest error rate in any county recount was .73%. Most counties found no change in their finally tally. The majority of the remaining counties had changes of fewer than ten ballots.

Because the margin is still less than 0.5%, the President can request a recount after certification of the results. That recount will be conducted by rescanning all paper ballots.

[Click here for the Risk Limiting Audit Report](#)

###

More News & Announcements

Secretary Raffensperger Praises County Election Director Nancy Boren, Dismisses Attacks from Dark Money Group

Secretary of State Trains Local Officials on New Voter Registration System

Federal Judge: Abrams Group Must Repay Costs to Georgia Taxpayers

Call for Special Election for State House, District 119

Call for Special Election for State Senate, District 11

Call for Special Election for State House, District 172

[VIEW ALL](#) ➤

[Office of Brad Raffensperger](#)

[News & Announcements](#)

[Privacy Policy](#)

[Security](#)



214 State Capitol
Atlanta, Georgia 30334

[Contact Us](#)

© 2023 Georgia Secretary of State

[TOP](#)